Applied Experimental Psychology

Applied Experimental Psychology

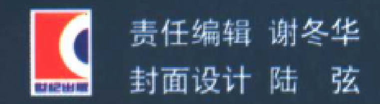
邓铸 编著

应用实验心理学

注重实验研究者思维方法的启发与训练，避免该课程沦为培训实验员的操作课程

总结国内近20年来实验研究方法的新成果，介绍最新、最有效、最常用的实验研究方法。

上海教育出版社
SHANGHAI
EDUCATIONAL
PUBLISHING
HOUSE



责任编辑 谢冬华
封面设计 陆 弦

Applied
Experimental
Psychology


ISBN 7-5444-0701-2
9 787544 407014 >
易文网：www.ewen.cc
定 价：48.00 元

江苏省哲学社会科学研究“十五”规划基金项目成果(04JYB010)
南京师范大学“十五”重点课程建设基金项目成果

Applied Experimental Psychology

■ 邓铸 编著

应用实验心理学

Applied Experimental Psychology

上海教育出版社
SHANGHAI EDUCATIONAL PUBLISHING HOUSE

图书在版编目(C I P)数据

应用实验心理学 / 邓铸编著. -上海:上海教育出版社,2006.6

ISBN 7-5444-0701-2

Ⅰ.应… Ⅱ.邓… Ⅲ.实验心理学 Ⅳ.B84

中国版本图书馆 CIP 数据核字(2006)第 072048 号

应用实验心理学

邓　铸　编著

上海世纪出版股份有限公司
上　海　教　育　出　版　社　出版发行

易文网:www.ewen.cc

(上海永福路 123 号　邮政编码:200031)

各地新华书店经销　太仓市印刷厂有限公司印刷

开本 787×1092　1/16　印张 33.5　插页 3

2006 年 6 月第 1 版　2006 年 6 月第 1 次印刷

印数 1—5,000 本

ISBN 7-5444-0701-2/B·0008　　定价:48.00 元

序

2006年春节前，一直惦记着去看望我的实验心理学的启蒙老师——河南师范大学的华克伦教授。1989年，我是先生的助教；1990年，先生是我的"助教"！那一年，我开始讲授实验心理学，教室后排坐着我那位白发"助教"！每次上课，站在讲台上，看着那张慈祥的面孔，心中充满感动，语气中也不知不觉地增添了许多激情，鼓动得不少学生开始热爱起心理学来。然后，我又被学生鼓动着，也跟着热爱起心理学来！十几年过去，懵懵懂懂地往前走着，其实靠的就是那一年的感动，那一年的鼓动！

不久前的一天，见到了阔别已久的华克伦先生和陈慧芳教授，两位相濡以沫的老师除了苍老许多外，性情还是那样豁达，谈吐中依然带着几分天真。我突然觉着有些惭愧，十几年中竟然未能成就一件可以呈给恩师的作品。

那些年，我讲这门课也不容易，到了讲台上，虽然也煞有介事，但肚里没货，总是很心虚。后到中国科学院心理研究所短期进修，又到陕西师范大学心理系学习。在欧阳仑、杨永明、孙昌识、方俊明、李殿凤等教授指导下系统地学习心理学课程。在西安，大嗓门的欧阳仑教授给了我不少鼓励！1995年，我有幸来到西南师范大学心理学系，追随我国著名心理学家黄希庭教授，在先生指导下从事人类认知的实验研究。在这里，我接受了最严格的学术训练，先生以他渊博的学识、严谨的作风、宽容的胸怀影响着弟子。那时学友们都羡慕我，因为常常在傍晚时分，窗外有一个带着浓重江浙口音的人在叫我，那是先生让我陪他散步，这是最让我感到受宠若惊的时刻！在我感觉中，陪先生散步就是先生为我"开小灶"，有了更多的机会感受大家风范，也因此有了对未来更多的希冀。

离开西南师大，回到实验心理学的课堂，感触已不同于从前。作为基础心理学的核心组成部分和心理学方法学课程，实验心理学在心理学专业人才培养中起着重要作用，因此这一课程的教学体系建设本身就是一个具有吸引力的课题。1999年，我考入我国心理学的重要发源地之一南京师范大学攻读博士学位，接受有不同思维风格的杨鑫辉、余嘉元、叶浩生、郭亨杰等心理学家的训练，先后得到过杨治良先生、沈德立先生、孟庆茂先生的指导。几年来，我承受着心理学实

验研究和理论研究的双重诱惑,也分明感到心理学和心理学家存在的分裂,即科学主义和人文主义阵营的分裂,这是国际性的。尽管这种分裂可以激励两个阵营的心理学家各自更勤奋地工作,但我认为两种文化的分裂可能是不必要的,也肯定是有害的。关于这一点,我还是有很多感动:车文博先生在南京师范大学做过精彩的演说,他以理论家的睿智,细细阐述心理学实证研究的重要性;杨鑫辉先生,这位中国心理学思想史专家,一直推动着心理学的技术学研究;我的导师余嘉元教授,勤奋、朴实,奔波于大学讲坛与企业工场之间,将联结主义的最新理论灵活地应用于人力资源管理和工程设计过程。所有这些都给予我受用不尽的启迪!郭本禹教授,在我学习与工作的几年中,给予我不少指导与帮助,特别是他使我有机会接触和翻译了格拉维特博士和福泽诺博士的《行为科学研究方法》①,让我对心理学实验研究的逻辑与方法有了更深入的理解。

叶浩生教授是我国心理学理论研究和历史研究的多产作家,同时又是南京师范大学心理学实验室建设和实验研究、应用研究的有力推动者。在短短的几年时间里,南京师范大学心理学科建成了"基础心理学实验室"、"认知神经科学实验室"(ERP)、"眼动仪实验室"(SMI)、"认知与情绪实验室"、"脑与行为实验室"、"语图分析实验室"、"E-Prime 实验室"……拥有了一系列现代化的心理学实验研究基地,组建起一支年富力强的实验研究学术梯队。在南京师范大学教科院,我分明感受到,心理学的理论研究与历史研究、心理学的实验研究与应用研究都在快速地发展,由此我知道了:心理学家个人可以做他擅长的研究,心理学家之间可以从事着完全不同的研究,但是作为心理学家共同体,必须能够包容各种不同取向的研究。

在学习的生涯中,承沐如此众多的关爱,何以回报?唯有认真地工作!教学中,我开始寻找能将心理学的理论研究和实证研究融合贯通的逻辑一致性,并对国内外不同版本的《实验心理学》教科书进行比较和研究。国内现有实验心理学教材是以杨治良先生的《基础实验心理学》(甘肃人民出版社)、《实验心理学》(浙江教育出版社)和朱滢先生的《实验心理学》(北京大学出版社)为代表的,这些教材的结构框架成为我国实验心理学教学体系的蓝本,在我国心理学专业人才培养中起到了非常重要的作用。近年来,国内不少学者开始探索新的教学体系,这是我国心理学人才培养规格和规模都迅速发展的形势所迫切要求的。除一系列

① 已由陕西师范大学出版社出版(2005)。

翻译类著作外(郭秀艳,邓铸,张奇等),还出版了不少新的《实验心理学》教材(孟庆茂,郭秀艳,金志成,欧阳文珍等)。这些翻译和编著的教材开始有了不同于前述经典教材的一些特点,特别是在内容体系的调整、国内外新研究成果的吸纳方面,都进行了不少尝试,体现出各位作者的学术勇气和渊博学识。

在十几年的教学过程中,我也在尝试不断地对教学内容进行补充和调整,并从以下三个方面进行实践探索:

第一,注重学生理性思维的训练。心理学是深入人们日常生活的学科,学习或不学习心理学的人似乎都能或多或少触及心理学知识,因此不少学习心理学的人对严格的方法学训练不以为然。在实验心理学的教学中,大多数学生的学习兴致都很高,但也有的只是因为觉得实验心理学“挺好玩的”。我也越来越多地感到,不少学生在学习中缺乏一种“研究意识”,甚至到了大学阶段,还会认为教材上讲的、杂志上发表的、学术权威口里说出的都是对的。我在教学中强调,对于书本知识、文献材料、大家演讲,起码要把握两条评判标准:一是事实本身的样子,二是科学研究自身应有的内在逻辑,这是最基本的理性思维。心理学的实验研究有自身的内在逻辑,它能帮助我们把握研究进程,尽量少犯错误。学院式的训练不是培养现代心理学专业人才的唯一途径,但却是提升培养规格和专业素养的重要途径之一。

第二,强调心理学实验研究基本程式的把握,特别是心理学实验设计方法的掌握。在实验心理学课程体系中,有一个核心的内容就是“实验设计”,它强调要根据课题的性质,选择和分派研究被试,选用实验设计模式,操纵或控制各种变量,整理和分析研究资料及数据。这种严格的要求至少有三方面的意义:首先,这些成型的实验设计模式在实验心理学长期发展中被证明是有效的、可靠的,它使得心理学的实验研究有序和能够克服研究者的个人偏见。其次,实验设计模式与数据分析方法具有对应关系,所以只有按照一些规范的模式设计实验进程,数据的分析才不会出现技术上的困难。要想避免在面对无序的数据资料时一筹莫展,就要在研究的开始严格遵循实验设计规则。最后,研究结果的公开、发表和交流,要求其研究过程的规范性,也就是说,研究进程的把握和研究成果的报告,都需要使用心理学家的“行话”,以便于交流和互相评判。

第三,拓展学生的观察视野和开展应用研究的技能。在注重经典实验范例选择与介绍的同时,引入了一系列新的研究范型和成功实验范例,还增加了两章心理学应用研究领域的实验方法的介绍,以拓展学生的观察视野。在这一过程

中,增加具有很强应用性质的教学实验,将基础实验研究和应用实验研究的技能与方法融通。特别是,在教材中介绍了“人因工程学”的基本概念、实验研究方法,这对于学生将学院式学习与社会生活联系起来具有重要意义。因为强调了应用性,所以决定将书名定为《应用实验心理学》。当然,强调“应用”还有另外一层意思,那就是强调全书介绍的方法的“可用性”。

在进行教学内容的调整中,紧密结合我国心理学教学实际,包括本科生教学和研究生教学的实际,既坚持方法学教材的严谨与规范,又力求语言的流畅与易读,尽力改变实验心理学学习的“苦行僧”方式。

本教材内容由四部分组成:

第一部分:心理学实验研究的基本问题,由导言、第一章至第三章组成。该部分主要讨论了心理学科学研究方法的主要特征、一般逻辑、心理学实验研究的基本过程和常用设计模式,是实验心理学教学的基础内容和核心内容。在这一部分,还以阅读材料的方式介绍了中国大陆实验心理学的发展历程、中国台湾实验心理学的奠基、心理学研究的生态化运动、内部效度与外部效度的平衡问题、研究文献资源及其检索、心理学研究中的伦理学问题,等等。这一部分要求学生深入理解,并能灵活应用,在培养科学精神的同时,能够结合实际课题编写出有一定质量的心理学实验研究方案。

第二部分:心理学实验研究的传统领域及其研究方法的新发展,由第四章至第九章组成。该部分主要介绍了心理物理学、反应时间、感知觉、记忆和思维等领域的实验研究方法及研究成果。编写中,兼顾了基础实验方法的介绍和经典实验的评介,尽可能体现这些领域实验研究方法的新发展,突出这些研究的基础性、应用性。通过提供阅读材料,将涉猎内容扩展到心理学基础研究在诸多实践领域中的应用,如反应时间测量的应用、现代心理学研究的反应时范型、听觉告警、汉字认知研究、认知方式研究、脑与记忆、记忆的应用、问题表征,等等,以增加教材的可读性,拓展学生的视野。

第三部分:心理学实验研究方法的应用,由第十章和第十一章组成。该部分主要讨论了发展与教育心理学的实验方法、人因工程学的学科体系与研究方法。人因工程学是一门亟待发展的新学科,它能有效地将学院式的实验心理学与工程设计紧密结合起来,促进实验心理学走向社会生活和各种实践领域。该部分还对发展心理学和教育心理学研究的特殊性进行了分析,展望了人因工程学的发展前景。

第四部分:实验心理学学习的辅助内容和扩展,由附录1至附录5组成。附录1是为常用研究类型的数据分析编写出可以直接运用的SPSS程序,这些程序可以帮助学生快速学会并使用SPSS软件分析实验数据和资料;附录2介绍了E-Prime实验生成系统在心理学实验中的应用,这是当前国内外最为流行的心理学实验生成软件系统;附录3给出了一系列常用的统计检验数字表,包括随机数字表、*PZO*转换表、t分布表、F分布表和χ^2分布表;附录4是术语表,列举出本书出现的专业术语等。

每一章均由“本章内容提要”、正文、“阅读材料”、“建议阅读文献”、“复习思考题”五部分组成,学生如能在使用教材的同时,研读章后提供的阅读文献,就可以极大地丰富课程学习内容,深化对课程内容的理解,起到事半功倍的效果。

写作过程中,蒋波博士给予了积极的帮助并完成了第六章“感觉的实验研究”。我的研究生姜子云、余仕华、陈庆荣等分别完成了第十章“发展与教育心理实验法”、第八章“记忆的实验研究”(一、二节)、附录2“E-Prime在心理学实验中的应用”的初稿。我的2004级和2005级研究生对书稿进行了研读,姜子云、陈庆荣、曾晓尤、余仕华、张婷婷、迟东娴、孙丽伟、杨富丽、易琳等为书稿修改提出了不少好建议。南京师范大学教科院2004级“应用心理学”专业本科班的全体同学,将本书的初稿作为内部教材试用,不少同学提出了非常好的修改意见和建议,为提高书稿的整体质量起到了明显作用,他们是章悦、陈晓、纪一川、张亮、吴黄旭、陈健申、姚学斌,等等。

写作中参考和引用了大量中外文献资料,恕未能在参考文献中全部列出,在此表示歉意和感谢!

在此书的写作与出版过程中,叶浩生教授给予了多方面的关心与指导,在此表示真诚的敬意和感谢!上海教育出版社的谢冬华先生,为此书的初期选题、编著体例、文稿的修改和定稿,做了艰辛和具有建设性的工作,在此表示诚挚谢意!

最后,请各位前辈、同仁和青年学生不吝赐教,不断提出宝贵意见和指正之言,以便我在今后的修订中采纳,弥补我学识之不足。先在此向您表示诚挚谢意!

邓　铸

2006年2月于随园

目　录

导言:实验心理学的发展与逻辑

第一章　心理学实验研究的过程

第二章　多因素心理实验设计

第三章　生态学方法与准实验设计

第四章　心理物理学

第五章　反应时间

第六章 感觉的实验研究

第七章 空间知觉的实验研究

第八章 记忆的实验研究

第九章　思维的实验研究

第十章 发展与教育心理实验法

第十一章　人因工程学研究

导言：
实验心理学的发展与逻辑

科学开始于人类的意识觉醒和生存斗争，人类意识中的好奇心是科学发展永不枯竭的内驱力。也许，人们可以通过“忌想”和“观察”来满足好奇心。忌想的“结果”是丰富和有趣的，观察的“发现”则引导人们走向无限的探索生涯，在经验中求证，又在经验中叩问世界的奥秘！实际上，希腊哲学中的唯理论和经验论最能反映人类认识世界的两条道路。但是唯理论可靠吗？我们只能说，其中的真知灼见永远闪耀着人类思想的光芒，但它却时常与观察到的现象不相吻合，人们无法不困惑犹疑。经验论可靠吗？我们一样不能绝对肯定，但却可以说：经验论更有助于积累实在的知识，并发展出一系列获取实在知识的方法和技术。就心理学而言，观察技术、测验技术、相关技术和实验技术等，在我们认识人的精神现象中都发挥着重要作用，而且都充分体现了心理学科学化的追求。

一、科学的方法与主观的研究者

（一）科学研究从观察开始

原始人类的认识活动从感觉开始，从对现象的感知开始。当自然力迫使原始人类走出森林，现成食物无法满足他们生存需要的时候，他们就开始种植和养殖。如何种植和养殖？天气对这些活动有什么影响？……诸如此类问题存在于原始劳动中，而这些问题在不断的现象观察中被解决，并成为人类早期生活中的实践经验。这些经验区别于动物的本能行为程序，可以被传授或被主动学习。这些经验的积累和结合逐渐超越了现象感知，形成人类最初的农业技术和畜牧技术，也成为科学的最初形态。

语言文字的出现和发展、劳动工具的发明和制造遂成为人类文明的标志，也是人类文明继续前进的促进因素。人类实践活动的范围和复杂程度越来越高，接触到的现象越来越多（包括开始注意到人与人之间的各种现象和关系），而文字可以很好地记录下他们觉得有趣的现象。当人类社会发展到可以使一

部分人从体力劳动中解放出来，专门去追索各种现象的前因后果时，他们就成为科学的真正先驱，其任务就是：查阅前人记录的各种现象、观察他所能观察到的现象（有时需要借助一些工具来进行观察，这些工具就成为科学的仪器）、尽其所能描述现象的特点和寻求它们之间的关系，这就是科学研究最基本的任务，即使是现在，科学研究发展得多么辉煌和精湛，它也没有脱离这一基本任务的范畴。

科学研究就是观察记录有关现象，揭示现象的本质特征及相互关系。

不过，今天的科学研究程序与最初情形相比有一个重要而明显的不同。今天的科学家都拥有某一学科或多个学科的基本理论知识，这就意味着他在还没有进行真正的科学研究之前就对有关现象及其关系有了比较深入的理解，在此基础上，他就可以重新审视已有的各种理论并就未知的规律提出假设，然后用观察到的现象进行检验。这种现代科学研究的程序使我们少走弯路，更有成效，但同时，科学家的或说整个人类的自以为是对研究结论可靠性的影响也变得更为微妙。归根结底，科学研究还是从感觉开始的，而且科学结论的正误还需回到实践中去检验。科学的最根本特征还是经验实证的。

（二）观察不总是可靠的

正像心理学研究后来证明的那样，从事科学研究的观察者持有许多经验，先入为主的信念会或多或少地影响甚至左右其观察活动（Greenwald, Pratkanis, Leippe, et al., 1986）。人类普遍存在着求证误区：有意无意地去寻求那些支持其假设或信念的现象，回避那些反驳其假设或信念的现象。此时此地的我又怎么见得没有犯这样的错误呢？即使如此，我们并不会停止对世界的探索和讨论，又总会不遗余力地为我们的各种理论寻求支持，因为我们总会对自己头脑中浮现的那一丝灵感非常自信。聪明的人们，有时对于自己的愚蠢竟那么熟视无睹，甚至还为之洋洋得意！

此类例子并不少见。14～16 世纪，欧洲战场上的医生们曾经习以为常地使用“滚油灼烧”止血技术，即当士兵在战场上受伤不得不截肢时，战地医生就会在手术台旁支起一口大锅，将食用油倒进去煮沸。等到手术截肢后尽快用滚油浇灼伤口以止血。当时，这一技术被认为非常有效而广为采用。然而，在一次战役中，仗打到一半，一位法国医生安布诺伊斯·佩厄尔就无油可用了，他为士兵做截肢手术后只好直接用止血带简单包扎伤口。可是，后来他发现，这些未使用“滚油灼烧”术的伤员的伤口愈合得更好（Levine & Parkinson, 1994）！这一欧

洲战场上流行300年的止血术从此才被废止。[①]我们还注意到在几乎所有的医院里,医生对病人的诊断和决定是否施行某种手术都显得过于自信和不容置疑。如果他们诊断后认为病人已无可救,病人却逐渐康复,医生会说:真是奇迹!如果为病人施行了某种治疗,病人却还是死去,就将其归结为身体太虚弱或是送来得太迟了,等等。医生从来不愿承认诊断的错误或某种治疗方案的无效性。

在行为科学研究中,研究者先入为主的信念对研究结果的影响比在自然科学研究中的影响更为严重和更难以发现,因为人的心理和行为绝不像自然现象的运动那样遵循严格的法则,它具有一定的不确定性,所以心理学发展中的争论更多。但是它绝不是无章可循,任何现象都是偶然性与必然性的统一,心理现象也绝不例外。既然我们能建立科学,那就毫无疑问地说,我们有办法克服个人偏见,从许许多多的偶然现象中发现必然性。

(三) 科学的研究方法

科学的建立,或说一门科学所以能独立存在,需要几个方面的基本条件。首先,它有自己的研究对象,即某一类独特的现象;其次,它是社会需要的,即它的研究成果有用,这可以保证它有自己的专业人员和追随者,有些人可以在这个科学领域中实现价值并得到社会的承认;第三,它必须有自己科学的方法学体系,这个方法学体系保证它能尽量避免研究者的个人偏见,广泛地收集与其研究对象有关的许多偶然表现,从中揭示这些偶然表现中的必然性。只有揭示出必然性,它才可能是有用的、可以指导人的实践。这里,我们单就科学方法来说,它需要具备以下特征:

1. 科学方法的经验性

如人们所知,当我们说科学是经验性的,就是说问题的答案是通过观察得到的。尽管预测性的答案或假设也许是通过其他方式得到的,但科学要求经验证实。根据共同的感觉,一个答案可能是明显的,而且也可能完全符合逻辑,这个领域中的专家也都支持它,但在它还没有得到经验证实的时候它就不是科学的。

科学方法包括结构化的或系统的观察。观察结构是由研究中使用的程序和技术决定的。更具体地说,观察的目的是为假设提供经验的验证,因此观察是结

---

① Levine, G. & Parkinson, S. (Eds.) (1994). *Experimental methods in psychology*. New Jersey: Lawrence Erlbaum Associates, 1—3.

构性的,以便它能够为假设要么提供明确的支持,要么提供明确的反驳。现在考虑:维生素C预防流行性感冒的作用有多大?

要解答这个问题,只询问人们是否定期服用维生素C、在某一季度里患过多少次感冒,显然是不够的,因为这样的观察没有结构性,无论得到什么样的回答,都不能为问题提供确定性答案。特别是,我们没有确定个体服用维生素C的剂量水平,也没有查验那些报告患病的患者实际上患的是某种流行性感冒还是肺炎或其他疾病,更没有考虑被调查者的年龄、总体健康水平和生活方式。我们也没有试图控制人们关于维生素C和感冒的信念可能对其回答的影响而带来的偏差,没有去对那些每天都服用一定剂量维生素C的人与那些服用冒牌药丸(安慰剂)的人进行比较……我们可以列出更多方面的问题来说明,这里的观察是无结构性的,所以也肯定得不到确定性结论。

在科学方法中,观察是系统的,它要在一系列特定条件下完成以便我们能够准确回答正致力于解决的问题,就是说,在完整的研究中,观察是结构性的才能检验我们关于这个世界运动方式的假设。如果你想知道维生素C是否能预防感冒,有一种方式可以使你的观察具有结构性从而得到问题的答案。本书介绍的许多方法都能做到这一点,会讨论到如何控制结构化研究以得到具有说服力的可能的解释。

2. 科学方法的公开性

科学方法是公开的。在此,我们的意思是说科学方法使得观察可以得到其他人的评估,特别是其他科学家的评估。在特定情况下,其他人应该能够按部就班地重复同样的观察过程,亲自复制这一观察。请注意,只有公开的观察才能够被重复,因此也只有公开的观察才是可核实的。

科学共同体通过将研究报告发表在科学杂志上而使得观察公开化,这一行动是非常重要的,因为任何私密化的事件是无法被别人复制或评估的。大部分杂志上发表的研究报告都已经得到研究者同行(同一领域中的其他科学家)的评估,这是为了保证其研究方法的严密性与适合度。为了保证研究中不存在明显缺陷,研究报告必须符合各种标准才能发表。当你阅读杂志上的文章时,有一个事情你要注意,那就是对研究方法描述的详细程度。研究报告一般都有一个独立的“方法”部分,它会很详细地描述研究了什么样的人或动物、运用了何种仪器和设备、进行了何种测量等。方法的详细程度应足以使他人能准确地复制同一研究以便对结果进行核实。复制和核查的理念是重要的,它们提供了对研究的

"检查和平衡"机制。①

就如我们将看到的,由于误差或偶然因素,存在多种可以导致错误结论的因素。有时研究者犯欺骗性错误而故意伪造或歪曲研究的结果。研究者仔细地审察和评估研究报告,对研究结果持一种怀疑的态度,直到更多的研究证实这些发现,这是非常重要的。通过复制研究和使之接受同行专家的审阅,我们有了重要的检查和平衡机制以防止错误。

3. *科学方法的结构性*

科学方法具有结构性,才能有效地保证研究的客观性。科学一直被称为"不动声色地搜索知识",意思是说研究者不让个人感情污染观察。研究中容易出现什么样的个人偏见或信念呢?通常,个人偏见来自特定的理论,研究者也许会试图寻找支持他/她自己的理论的证据。因为研究者一般都是在检验理论,他/她会对研究结果有一种期望。在有些情况下,期望会微妙地影响研究发现。

科学方法具有结构性,它就能在相当程度上制约研究者个人偏见对资料搜集、资料整理和做出结论的影响,以保证在收集全面真实资料的基础上正确解释现象间的关系。研究的客观性绝对不是单靠实事求是的科学态度就能保证的。比如说,要研究智力的发展水平受到什么因素的影响,我们就要考虑在一定数量的儿童当中尽可能全面地收集可能与儿童智力发展有关的资料,像年龄、性别、生理成熟水平(假设其中包含体重,甚至手掌大小)、父母职业、成长环境等等,然后你可能发现儿童智力发展水平与其手掌大小存在显著的正相关,即有明显的正向共变关系。在现象感知的水平上,我们似乎可以确信:人的手掌越大其智力发展水平越高。显然,这个结论是荒谬的。如果我们掌握了足够的方法,而且按照这些方法进行研究,就不大可能得出这样的错误结论。要想探明手掌大小与智力水平是否有关,自然有相应的方法来解决,即控制智力和手掌以外的变量,你就会发现在同一年龄阶段、同一性别、相似的生活环境下,智力与手掌大小的相关就不存在了。

还比如,我们要用再认方法比较两个大学生对同一单词组的学会程度,可以把这些单词与一定数量的没有学过的单词混淆在一起,分别让两人再认。结果甲认出学过单词的 80%,乙认出 65%,那么是否就可以说甲比乙学会的程度好

① [美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著,邓铸、姜子云、蒋小慧等译:《行为科学研究方法》,陕西师范大学出版社 2005 年版,第 14 页。

呢？不能确定，因为甲乙两人在再认测验时可能有不同的自信水平，而自信水平又对再认结果产生影响。因此，为了得出可靠结论，必须使用信号检测论的方法处理数据，排除自信水平的影响而真正看到他们对单词组的学会程度。

在一个科学领域中，方法的不断完善和可靠是其学科科学化的重要特征。

4. **科学方法的操作性**

科学方法应可以传授、易于掌握，它能够成为专业人员从事研究时广泛使用的工具，又可以成为一般读者理解某一学科基本理论知识的钥匙，同时它也能起到沟通理论研究与应用研究的作用。就拿现在心理学研究中使用得越来越多的因素型实验设计来说，其思路明晰、适应面广，实验控制和操作变得简单易行，先进的统计分析技术又可以很好地把交织在一起的各个变量对人的心理行为的影响分离开来，研究结果也容易解释。

就心理学领域来说，会经常听到一些抱怨：心理学的专业刊物很难读懂，学习了心理学却不知道怎样利用它去解决实际问题，我想做一篇论文却不知从何入手，等等。这里缺乏的就是一种方法学的训练。进行方法学的系统训练以后，你就会发现国内最难读懂的《心理学报》、《心理科学》和《心理发展与教育》等刊物上的研究报告都是按固定格式撰写的，从中你很容易搞清楚：与此有关的前人的研究如何，作者提出了什么样的假设，为检验假设他采用了何种研究工具和设计方案，研究的对象是什么样的人或动物，实验操作的详细过程是怎样的，收集到了哪些类型的资料，最后采用了何种资料处理与分析方法，以及得到了什么样的结论，等等。如此一来，心理学的研究报告就容易读懂了。当然，在阅读时会遇到一些基本概念和基本理论理解上的困难，可是如果你只是选择你关注的某一领域的论文来读，这就又不构成障碍了。怎样应用心理学的研究成果呢？方法的训练对此也有帮助，因为把基本理论应用到实际工作中，开始总是带有尝试性或试验性的，而试验的方法往往与研究方法没有本质区别，只是被使用的面不一样，对过程的控制程度不一样罢了。

简单地说，科学研究的方法不能太神秘化（神秘化的方法往往是学术骗子使用的），要易于理解、易于掌握、易于推广。

5. **科学方法的系统性**

一门科学的各种研究方法应构成一个系统，从横向看，这个系统由多种方法构成，多种方法之间互相弥补、互相检验；从纵向看，它由哲学层次的方法论、规范研究的基本原则、收集整理资料并进行推断的操作技术构成。一种具体的研

究方法也是一个系统,它是由一系列操作步骤构成的一个完整研究过程。不能纳入到这样的系统的研究方法就不能说是科学的方法。在心理学中,一般不把精神分析的方法归入科学方法的体系中,因为它的结果具有相当的不确定性,其结论也往往难以得到其他方法的检验。

再回到开始,我们强调科学研究开始于现象观察,但我们又说,观察未必总是可靠的,这一矛盾如何解决?请记住:科学方法是一个系统!心理学研究的常见历程是:现象观察得到描述性的结果,相关分析显示变量间存在的关联,然而要确定变量间关联的性质就需要实验技术,所以说,观察技术、相关技术和实验技术是一脉相承的。此外,科学研究还存在另一条道路:理论或观念导致预测或假设,而假设应是可拒绝或可接受的,因此是可检验的,于是就有了观察、相关或实验研究的任务——检验假设。

(四)研究者的主观性

因为研究者的主观性,研究方法就要具有科学性。科学研究总是由科学家来完成的,科学方法是客观的,但科学家总是具有主观能动性的,既有真知灼见,也有个人偏见。明白这一点至少有四方面的好处:(1)科学家的研究需要遵循某些规则,使用科学的方法,以防止个人偏见对科学结论的不利影响。(2)科学家之间既需要批评,又需要宽容。需要批评,就要将研究过程和研究结果公开,接受同行专家的评判和修正;需要宽容,个人偏见或信念导致的错误,与科学欺诈有本质不同,允许存在,也允许研究者去修正和发展。(3)任何一个科学家都不要指望以自己个人的才智承担一个科学领域发展的重任,而是要自觉地投入到科学家的共同体当中去。任何科学家都可以期望但不必强求自己通晓一个领域中的所有技术,而应尽可能地发展自己的研究技能、用己所长对科学共同体有所贡献。(4)任何学术权威都不能将自己的观点强加于他人,尤其是不能强加于年轻的学者,因为这样做的结果不仅会打压新生代的创造力,也会暴露自己的心虚。反之,青年学生、初涉研究的学者不要总是以为:权威者的演讲、教材上的洋洋文字、学术期刊上的研究报告都是对的。我的建议是,当聆听演讲或阅读书刊上的文字时,即使你一时无法对之进行全面检验,你至少可以看他/她的观点与你的经验是否吻合,这样想就容易引出新课题。

由于不同的知识经验和思维方式,科学家提出的假设前提、选择的研究手段、对结果所作的解释,都必然带有主观性。但是科学家的主观性可以得到多方面的制约和监督,比如遵循确定的研究程序、参与科学家群体的讨论、接受其他

有关研究的检验和其他科学家的批评，等等。

总之，科学的研究方法既是一个学科发展的前提，也是衡量一个学科成熟水平的标志；既是学习一门科学知识的有效途径，也是从事科学研究的基本工具；既是一个科学家群体联结和沟通的纽带，也是科学家个人偏见对学科发展不利影响的制约机制。《应用实验心理学》要介绍的就是这样一种科学的心理学研究方法。

二、实验心理学的建立与发展

实验心理学不像实验物理学、实验生物学等那样容易让人接受，因为它研究的对象是人的精神现象、内隐现象。在漫长的历史中，人们感到主观世界的变化无法预测、经验实证显得多余。18 世纪中叶以前，心理学尚未成为一门学科，人们对灵魂现象的好奇心在相当程度上还依赖“忌想”来满足。不过，从 18 世纪中叶开始，哲学开始解放这一科学研究的禁区：人是机器、人是动物、人的灵魂和肉体可以分离……这些信念开始深入人心，科学方法也开始深入“人心”了。

（一）实验心理学的酝酿

心理现象成为独立的科学研究对象尽管很迟，但人类对自身精神世界的探索却很早。在中国和希腊古代的哲学与医学典籍中有许多关于心灵或灵魂的学问，这些学问没有明确的研究核心和科学的逻辑体系，还不能构成一个独立的科学研究领域。到中世纪前后，欧洲自然科学繁荣，一些自然科学家注意到人的心理问题，并获得一些关于神经系统、感官生理和感知觉的知识，但是他们仍然固守一个信念——实验的方法不能用于心理现象的研究，此时关于心理现象的一些思想认识被称为哲学心理学或思辨心理学。心理现象还只是在哲学中附带加以研究。

从 18 世纪开始，欧洲哲学心理学的研究头绪纷繁，但各种观点互相矛盾、冲突，可见关于人的精神现象的问题是多么困扰人类自身。如果只是做一个轮廓上的描述，我们可以从洛克和笛卡儿开始，哲学家和科学家都表现出一种探索人类心理规律的急切心情，而且在他们的哲学著作中，随时都可以发现有关于感知、情绪、意识等方面的论述。后来的哲学家在对洛克、笛卡儿的唯心主义的批判性继承中，逐步建立起一种有利于心理学诞生的科学氛围，提出人是一种动物、人是一架机器、人的心理依赖人的生理活动、人脑是意识的器官、感觉是一切

意识的开端,等等。我们姑且不说这些观点是否正确,但它确实为将自然科学的手段和方法运用于心理现象的研究做了准备。哲学中的唯心主义、机械主义、经验主义和联想主义对实验心理学的诞生都起到过重要作用。没有这些思想和资料的积累,人们就不能摆脱宗教神学,就不能摆脱上帝,也就不能把人看成一架机器,一些自然科学的研究成果就不可能成为科学心理学的研究基础。如果人体不能分解、心理不能分解,那么人体就是一个谜,人的心理更是一个谜,人的行为只有上帝可以解释,心理学就没有了,心理学家就没有了。因此,哲学心理学是实验心理学诞生的必要的思想准备。科学家的成果启发着哲学家,哲学家的思想武装了生理学家,生理学家从感官生理自然接触到了感知问题,由此开始对心灵进行系统的实验探索,这就是实验心理学诞生的一般逻辑,所以第一代实验心理学家大都是学医出身的生理学家,他们同时又谙熟 18 世纪的欧洲哲学。

从 18 世纪中叶到 19 世纪中叶,可称为实验心理学的酝酿时期。此阶段,德国由于社会生产力和自然科学的迅速发展,在与心理学有密切关系的生理学方面一跃而居世界之先。生理学家缪勒(Johnnes Peter Müller, 1801—1858)及其弟子赫尔姆霍兹(Hermann von Helmholtz, 1821—1894)共同建立了"感官生理学"。解剖学家韦伯(Ernst Heinrich Weber, 1795—1878)根据多年的研究成果,获得了一些关于人的感觉的规律性认识,即后人所称的"韦伯定律"(Weber's Law)——$K=\Delta I/I_0$。与此同时,德国莱比锡大学的物理学教授费希纳(Gustav Theodor Fechner, 1801—1887)放弃了物理学,开始寻求精神与物质统一的规律。

费希纳生于德国乡村的一个牧师家庭,16 岁进入莱比锡大学学医,1822 年大学毕业时学习兴趣又转向物理学,1834 年被聘为莱比锡大学物理学教授。他很快成为一名才华出众的年轻物理学家,同时对心理学有研究兴趣。他长期对视觉后象进行观察并因此患上眼疾,1839 年辞去物理学讲席,卧病达 11 年之久。他在受尽病痛折磨的同时,也得以有时间思考灵魂问题。他开始用一种富于诗意的神秘眼光考察世界。他认为凡物有组织就有生命,有生命就有灵魂,精神与物质统一于灵魂。费希纳后来将主要精力集中于用科学方法证明他的这种泛灵论哲学。经过长期研究,他于 1860 年出版了著名的《心理物理学纲要》,此书的出版受到广泛重视,也招致诸多批判,于是他又用生命的最后 11 年时间研究心理物理学(psychophysics),竭力为心理物理学辩护。事实上,费希纳的《心理物理学纲要》在心理学史上具有开创性的和深远的意义,以致后来许多心理学家认为心理学诞生于 1860 年。

费希纳的心理物理学是关于身心之间或外界刺激量与内部活动心理量之间函数关系或依存关系的严密科学,是介于物理学与心理学之间的一门独立学科。费希纳坚信心理量是可以测量的,再加上他具有坚实的物理学和数学基础,对感觉阈限进行了深入的实验研究和精密的数学论证,由韦伯定律引申出一个感觉强度与刺激强度的所谓心物关系的对数定律:$S=K\log R$,依照这一定律,心理量的算术级数与刺激的几何级数成正比。费希纳还制订了系统精巧的心理物理法,这就是沿用至今的最小变化法、恒定刺激法和平均差误法。这是心理实验方法的第一次概括和总结,费希纳为此付出了艰苦的努力,比如他使用恒定刺激法测量重量的差别阈限,于1855年到1859年间共进行了6.7万次比较。费希纳的心理物理法与以前的心理实验相比具有两个特点:(1)利用了专门为研究心理物理学而制订的实验方法;(2)为了尽可能得到准确可靠的结果,可以多次重复同一实验,并对结果进行数学处理。这样的两个特点,在心理实验的发展中,标志着心理实验的方法向前迈出了重要一步,心理物理学方法的制订意味着心理实验"心理学化"的开始,意味着从比较草率的不精确的实验逐渐向着萌芽的实验心理学的过渡。①

在这一阶段,除费希纳的系统实验研究之外,还有其他学者进行的心理实验研究,但都比较简单和分散,总的来看,此阶段的心理实验具有以下特点:(1)运用的方法类似于某些物理学和生理学的方法,研究的问题多局限在某些简单的心理现象的量的方面,如视敏度、正后象的延续时间、差别阈限的测定、反应时间的测定等等;(2)实验的技术简单、有控制的重复实验及对实验结果进行数学处理总体上还比较欠缺;(3)被试的自我观察和陈述都还带有比较初级的性质,如听见或听不见某一个声音刺激、观察到或观察不到两个刺激的强度差别等。②

但不管怎么说,这个时候,一门采用科学方法系统探索人的灵魂世界的科学呼之欲出!

(二) 实验心理学的建立与心理学家共同体

实验心理学的建立开始于19世纪60年代,并和冯特的名字联系在一起。费希纳的心理物理学出版以后,更加坚定了冯特开创一个新的科学领域的决心。他总结和继承了前人的研究成果,于1862年出版了《感官知觉理论贡献》论文集。在此书的导言中,冯特首次使用了"实验心理学"的概念。1879年,冯特在

①② 杨治良:《基础实验心理学》,甘肃人民出版社1988年版,第2—3页。

莱比锡大学创建心理实验室,即标志着科学心理学的诞生。

> 心理学的诞生是件默不出声的琐事,未曾有一丝张扬。这天,在莱比锡大学一栋叫做孔维特(寄宿性的招待所)的破旧建筑物三楼的一间小屋子里,一位中年教授和两位年轻人正张罗着一些器具准备实验。他们在一张桌子上装了一台微时测定器(一种铜制的,像一座钟一样的机械装置,上面吊着一个重物,还有两个圆盘)、"发声器"(一个金属架子,上面升起一只长臂,有只球会从这里落下来,掉在一个平台上)和报务员的发报键、电池及一台变阻器。然后,他们把这五件东西用线连接起来,这套电路比今天开始电气培训的初学者用的那套不会复杂到哪里去。
>
> 这三位是威尔汉姆·冯特教授,一位47岁的男人,脸长长的,简朴的装束、满脸浓密的胡须;他的两位年轻学生,马克斯·弗里德里奇,德国人及G·斯坦利·霍尔,美国人。这套摆设是为弗里德里奇做的,他要用这套东西收集博士论文所需的数据。他的博士论文题目是"知觉的长度"——即受试者感知到他已经听到球落在平台上的时候,到他按动发报键之间的时间。没有记载写明那天是谁负责让球落下、谁坐在发报键跟前,可是,随着那只球"砰"的一声落在平台上、随着发报键"喀"的一响、随着微时测定器记录下所耗费的时间,现代心理学的时代就到来了。
>
> ——摘自《心理学的故事》(李斯,1999)①

威尔汉姆·冯特(Wilhelm Wundt)于1832年出生于德国一牧师家庭,十几年后成长为一名热情好学的青年,并进入大学学医,以后连续受到多位著名生理学家的培养。1874年,冯特开始研究心理学体系,1875年到莱比锡大学任哲学教授,于1879年建立了世界上第一个比较正规的心理实验室。该实验室一经建立,就吸引了世界上其他国家的一些青年学生,这些青年学生中的许多人后来成为著名的心理学家,因此我们可以认为,在冯特的周围,一个心理学家的共同体正在形成。此后,他的名声日重一日,1889年被任命为莱比锡大学的校长,1920年去世。

冯特是一位知识渊博的学者,一生著述很多,历史功绩主要在于他在心理学上的开创性工作。他在总结前人研究工作的成果之后,顺应时代对心理学的要

① [美]墨顿·亨特著,李斯译:《心理学的故事》,海南出版社1999年版。

求，建立了世界上第一个正式的心理实验室和比较完整的实验心理学体系，强调实验方法是心理学的主要方法，并在他建立的心理实验室里对感知觉、反应时间、注意、情绪和联想等心理过程进行了大量研究；同时创办了《哲学研究》，用来发表研究成果，使心理学终于从哲学的附庸地位中解放出来成为一门独立的科学。在此过程中，他培养了世界上第一代、第二代心理学家。

冯特的心理实验就是系统的自我观察(self-observation method)，或称为内省法(introspective method)，一切实验手段只是观察的辅助手段。心理学的研究对象是纯粹的“直接经验”，即通过自我观察而直接感觉到的“经验”。“经验”是由许多心理元素构成的。他希望通过“内省”把“经验”分解为简单的心理元素，如感觉和感情。冯特认为，实验心理学的主要任务是在严格控制的自我观察的帮助下精确地分析个体经验，即心理功能只能分解成简单的感觉成分才可以放到实验室里去研究，高级的心理过程是不能用实验方法来研究的。

冯特的学生铁钦纳(Edward Bradford Titchener, 1867—1927)发展了他的“内省实验法”，形成了构造主义心理学(structuralism psychology)。铁钦纳于1867年生于英国，入牛津大学学习哲学和生理学，后慕名到莱比锡大学追随冯特从事心理实验研究，成为冯特最得意的学生。1892年入美国康奈尔大学任教，传播冯特的心理学，并完成构造主义心理学的建立。铁钦纳坚决反对美国当时盛行的机能主义和行为主义倾向，不主张心理学研究心理或意识的生活功能，不同意采用动物、儿童、病人等作为心理学的研究对象，不主张使用心理测验，不承认潜意识观念，只强调采用内省法研究经验。认为经验构成意识，意识包括三种元素性状态：一为感觉，属于知觉之元素；二为想象，属于观念之元素；三为情感，属于情绪之元素。心理学研究的目的就是了解这些元素如何构成人的经验。显然，构造主义心理学与当时美国的心理学潮流及社会现实相脱节，受到同时代其他心理学流派的攻击，尽管如此，这一学派由于冯特的绝对权威和铁钦纳的雄辩口才，依然作为美国当时心理学的主流而存在，该学派在感觉等研究中取得重要成果的同时，也成为心理学发展的一种束缚力量，只是在1927年铁钦纳去世后，这一学派也宣告结束。

应该说，19世纪与20世纪交替之际是实验心理学建立和发展的时期。冯特的心理实验室成为当时世界心理学的中心，为以后的心理科学培养了第一代和第二代心理学家。但冯特的心理学也有许多缺点，他本人是唯心主义者，但有时又是唯物主义者，因此列宁评价他是“抱着混乱的唯心主义观点”，“像只老麻

雀”,经常在唯心主义和唯物主义之间跳来跳去。①冯特的实验心理学不可避免地产生许多矛盾,存在唯心主义形而上学的片面性,其后期成为心理学进一步发展的束缚力量。不过冯特作为新心理学的缔造者,他的历史功绩永远不会磨灭。美国心理学家霍尔(Granville Stanley Hall, 1844—1924)于 1921 年在哥伦比亚大学讲演时说:“冯特到任何时候都将作为伟大的里程碑而永垂不朽!”②美国心理学史专家黎黑则非常中肯地评价说:“冯特对心理学的长远重要性在于对社会习俗的影响,因为正是他开创了一个为社会所承认的独立学科,也为从事这一学科的人们创造了一种社会角色。”③

但有一点需要指出,一门科学的建立绝非个人可以左右的,它有人类知识发展内在机制的支配。在冯特时代,实验的心理学也在孔维特小楼之外酝酿发展着,典型的当数艾宾浩斯的记忆实验。不过,艾宾浩斯是在受了费希纳《心理物理学纲要》的启发之后开始记忆实验研究的。看来,这个时候,建立实验的心理学已是必然的科学运动了。那么,在冯特的实验室之外出现了什么样的心理学实验研究呢?主要有因素型实验、测验式实验、动物心理实验、儿童心理实验等。④

1. **因素型实验**

也称函数型实验,它是从艾宾浩斯(Hermann Ebbinghaus, 1850—1909)研究记忆开始的。这种实验的任务不是精确地分析认识过程,而是试图找出一定现象的原因,或者阐明两个变量间的函数关系。在这种实验中,起决定作用的不是被试的自我观察,而是被试在控制条件的实验当中完成某项作业的情况,如研究记忆、技能等的实验。

2. **测验式实验**

为了获取个体心理特征的材料而对之进行的测验,就是测验式实验,它包括直接测验和间接测验两类。直接测验就是对个体的个别简单心理机能的测量,比如个别感官的感受性、记忆能力、反应速度等;间接测验就是对个体某些高级心理机能或能力的测验,比如智力发展水平测验,这是通过测量一系列个别能力(感觉能力、记忆能力、解决问题能力等),然后以各种能力测验的综合结果作为个体综合智力的发展水平。

① 列宁:《列宁全集》(第 14 卷),人民出版社 1985 年版。
② 转引自黄珉珉:《现代西方心理学十大学派》,安徽人民出版社 1990 年版。
③ 高申春:《冯特心理学遗产的历史重估》,《心理学探新》,2002(1):3—7。
④ 杨治良:《实验心理学》,浙江教育出版社 1998 年版,第 32 页。

3. 动物心理实验

在19世纪末期,实验室中的动物心理实验逐渐发展起来。动物心理实验与正常成人心理实验相比,其最大的区别是动物实验中不可能给予被试指示语,也不能询问其内心体验。为了驱使动物去进行一定的活动,就要利用它的自然欲望,例如饥饿的动物获得食物的欲望、关在笼子里的动物获得自由的欲望等。因为不能询问它们在实验中的体验,许多现代心理学家在动物实验中特别注意观察和记录动物的行为及内部器官、系统的生理变化。

4. 儿童心理实验

儿童心理实验出现在动物心理实验之后,它在许多方面不同于成人的实验。用来进行实验的儿童的年龄越小,其与成人心理实验的差别就越大。幼小的儿童和动物一样,实验中不能有效地对其使用指示语,不能询问其内心体验,不能引起其复杂的活动。儿童心理实验的特殊困难在于儿童心理的快速发展。由于儿童心理的持续快速发展,在持续时间较长的实验中,研究者就很难确定实验的最终结果由实验条件制约到什么程度,由实验中的身心成熟制约到什么程度。儿童心理实验的关键是儿童对实验者的信任、对实验有兴趣,这往往需要赋予实验以游戏的性质。

(三) 分疆割据的时代:完形主义和行为主义

20世纪前半叶,是心理学继构造主义之后第一个学派林立的时代,这些学派主要包括机能主义心理学、完形主义心理学、行为主义心理学、精神分析心理学等等。这其中以实验为主要研究方法的心理学流派是完形主义和行为主义。

构造主义心理学后期,德国政治、经济、科学中一贯的整体观念,特别是胡塞尔(Edmund Husserl, 1859—1938)的现象学哲学推动韦特海默(Max Wertheimer, 1880—1943)、考夫卡(Kurt Koffka,1886—1941)和苛勒(Wolfgang Köhler, 1887—1967)等人创立了完形主义心理学(Gestalt psychology),音译为"格式塔心理学"。其基本观点是:心理现象都是有组织的、不可分割的整体,即完形。人和动物都可以通过内部的心理完形发现外部刺激情境中的完形,于是发生知觉,并解决问题。完形心理学反对任何形式的行为分析和心理分析,它的基本概念是"心理场"(psychological field),研究的主要问题是知觉和问题解决。问题解决方面最著名的是猴子或猩猩解决问题的实验,苛勒由此得出"顿悟学习理论"(insight learning theory),而在知觉方面的许多研究成果都保留在近期出版的心理学教科书中。格式塔心理学反对把现象分解,强调对现象完形进行整体观察,

而不是构造心理学的元素分解。它同时也反对行为主义对意识的否定，格式塔心理学的许多实验可以归为自然实验法或结构化观察法。

行为主义心理学(behavioristic psychology)是现代西方心理学的主要流派之一，对西方心理学影响很大。它是由 20 世纪初期在同构造心理学的斗争中由机能主义的极端化改造而来的，发起人首推华生(John Broadus Watson, 1878—1958)。华生成长于美国机能主义的策源地芝加哥大学，1903 年在芝加哥大学获得哲学博士学位，后留校任讲师至 1908 年，期间他经常与机能主义代表人物杜威(John Dewey, 1859—1952)、安吉尔(James Rowland Angell, 1867—1949)来往。1908 年他来到霍普金斯大学，1913 年发表了他的第一篇论文——《行为主义者眼界中的心理学》，这被看作是行为主义者的宣言性著作。随后他在《行为：比较心理学导论》(1914)中称："行为主义是唯一彻底而合乎逻辑的机能主义。"[①]1919 年，他通过《行为主义的心理学》对行为主义的观点做了全面阐述，并于 1924 年出版了一本通俗的《行为主义》，对他的学说进行宣传，行为主义从此就成为美国心理学的主流。

美国是实用主义哲学盛行的国家，他们把达尔文的动物"适者生存"论应用于解释人的行为。在政治、经济方面，他们强调心理的功能是适应竞争激烈的社会环境，这便是机能主义心理学，再往前一步就是华生的行为主义心理学。华生否认心理意识的作用，认为如把意识的研究作为心理学的对象，心理学就永远不能跻身于科学之林，所以有必要放弃以心理或意识为自己的对象，去研究有机体应付环境的全部活动——行为。[②]华生的观点迎合了美国当时的社会哲学，为积累资本，他们强调效率和秩序。当时，美国资本主义已经进入垄断阶段，行为主义把充分发挥人的潜力来增进生产效率、最大限度地提高利润、最稳定地维持社会秩序作为研究人的总目的。在行为主义者看来，生产效率是直接通过身体动作的效率来实现的，要提高生产效率就要提高动作效率，而维持社会秩序则在于人们的行为符合规范。心理学应该探索行为的规律，以使人的行为更有效率、更符合秩序。可见，行为主义是完全的机械唯物主义和实用主义，它的代表公式是：$R=f(S)$，其中 S 代表刺激(stimulus)，R 代表反应(response)。行为主义的研究方法就是在心理学实验室里严格控制刺激条件，观察记录人或动物的行为反应，完全放弃冯特的内省法。行为主义强调对人或

① Watson, J. B. (1914). *Behavior: an introduction to comparative psychology*. New York: Holt.

② 车文博：《西方心理学史》，浙江教育出版社 1998 年版，第 368 页。

动物的行为进行实在观察,所以实际上为心理学积累了大量实证研究资料,并对企业管理及其他经济活动产生了非常重要的影响。

华生的行为主义心理学不谈人的意识,较为极端,所以在20世纪30年代初,托尔曼(Edward Chase Tolman, 1886—1959)等人对其进行改造,创立了新行为主义心理学派,在 *S*—*R* 之间加上了机体变量 *O*(organism variable)或叫中介变量,使之成为 *S*—*O*—*R*,实验心理学的方程式就成了:$R=f(S, O)$。这一公式实际上是重新承认了意识的存在和意识的作用,基本上反映出外界刺激与内在心理活动之间相互作用的一般模式,是至今仍有重要意义的实验心理学方程式。

尽管行为主义从一开始就忽视心理意识的存在和作用,但在具体研究中,他们无法避开内在心理活动的影响,所以客观上揭示出许多有价值的心理活动规律。该学派的大量研究极大地丰富和发展了心理学的理论,它的许多经典实验在后来的实验心理学教科书中占有很大篇幅,特别是随着巴甫洛夫(Иван Петрович Павлов, 1849—1936)和斯金纳(Burrhus Frederick Skinner, 1904—1990)反射理论的建立,使整个心理学在管理、教育和社会服务方面具有广泛的应用价值,而具有更强的生命力。

实验心理学在这一阶段是积累实证资料的过程,同时它的方法也在不断地复杂化,精密仪器的制造使实验条件控制严密,对被试反应的记录也达到相当的精确度,对数据作系统分析的方法使实验心理学趋于成熟。

(四)认知研究的回归:对象与技术的扩展

"心理学流派的时代在1940年左右开始衰落了,并且这种严格划分心理学领域的做法也不再有可取之处。"①事实上,实验的心理学已经逐渐走出学派的疆域,几乎渗透到心理学研究的所有方面,也几乎渗透到人类的各个生活领域,其研究对象包括心理学所有可能的研究对象。心理实验研究技术的发展更是让其越来越走到科学世界的前沿。认知领域的多学科整合终于可以让人们对心理学研究有更多期望——揭开"智能"堡垒之盖。这一时代是从认知心理学(cognitive psychology)开始的。

1. 打开"黑箱子"

认知心理学是以信息加工观点为核心的心理学,兴起于20世纪50年代,到

① B. H. 坎特威茨、H. L. 罗迪格、D. G. 埃尔姆斯著,郭秀艳等译:《实验心理学》,华东师范大学出版社2001年版,第566页。

70年代成为美国心理学的一个主要方向，今天它已经成为国际心理学的潮流和前沿。认知心理学反对局限于研究孤立的、外观的、可观察的行为，而致力于了解人们内心进行的活动。这种活动就是传统心理学中称为认识过程的活动，就是全部信息加工活动。它的基本模型是：

信号输入⟶ 信息加工 ⟶行为输出

实验中控制信号输入，观察记录被试的行为输出，然后据此推断内部信息加工过程，这就是认知心理学的思路。认知心理学采用的方法是既重视实验室法，也重视主观经验的报告，更多的采用反应时间(reaction time，简称 RT)的测定法来建立内部信息加工模型，把实验心理学推进到一个很高的水平，使我们逐渐洞悉了人脑这一"黑箱"内的活动过程。这一发展也是实验心理学不断吸取现代科技成果的结果。

当前，认知心理学向着更精细化的方向发展，它广泛采用计算机技术，使刺激呈现、变量控制、反应记录和数据处理都达到非常精确的地步，而且通过模型建立、假设检验等方式，逐渐揭示出人脑内部的信息加工过程，其研究结果具有很强的理论价值和应用价值。人工智能、教学心理是认知心理学中两个重要的应用研究领域，前者可使心理学的研究成果为高技术产业服务、为金融服务、为大众日常生活服务，后者为心理学在教学中的应用提供丰富的理论基础和方法。如70年代开始的元认知(metacognition)研究就是直接与青少年的智能训练及提高学习效率有关的。

2. 认知科学时代的综合

实验心理学是科学大家族中的晚生代，然而今天逐渐走向成熟，开始与古老但依然繁荣的学科融合，这是它开始承担更大使命的时代的到来！

心理学家已不满足于对"黑箱"内事件的间接发现，也不满足于"准科学家"地位，他们着手将众多人文的、自然的科学成果进行综合，并引入各种大型、精密、高智能化的研究手段，试图直接发现大脑内部的活动。"美国国会曾断言20世纪90年代是大脑时代。认知心理学家已经热切地抓住神经科学以扩大我们对认知功能的理解。"①现在看来，20世纪90年代不过是大脑时代的动员期。进

① B. H. 坎特威茨、H. L. 罗迪格、D. G. 埃尔姆斯著，郭秀艳等译：《实验心理学》，华东师范大学出版社2001年版，第567页。

入21世纪后,大脑断层扫描照相、事件相关电位(ERP)记录、高速眼动追踪等技术开始逐渐在心理学的实验研究中得到推广。除此之外,我们依然不要忘记,心理现象始终不是物理现象,它不是靠力学定律解释得了的,其与物理现象最本质的区别在于文化因素。忽略文化因素,实验心理学就走向了极端,反而不能揭示心理生活的规律。现代认知科学不仅走向了认知神经科学的前沿,而且正在实现心理学、生物学、数学、物理学、计算机科学、语言学、文化人类学、哲学的大综合。真正的实验心理学家不会采用完全的自然科学方法和手段,他必须遵循实验心理学自身的研究逻辑。

心理实验在心理学科学化的道路上不断发展,其一般性变化主要有以下几方面:①(1)为了保证实验条件较完全的控制,制订和运用了心理实验研究的一般方法,如平行组——实验组和控制组的方法、平衡程度的方法、轮回的方法等;(2)制订和应用了实验研究的特殊方法,如记忆研究方法、注意研究方法、条件反射研究方法、情绪研究方法等;(3)在实验中制造和运用了或多或少的复杂仪器,广泛运用了最新技术成果——照相、录音、大脑断层扫描和计算机技术等;(4)在实验中运用了研究的统计方法,保证了在实验中可以同时应用两种或几种独立的变量。而且这些统计的完成可以依靠计算机,当前使用最多的统计处理软件SPSS可以对大批数据进行迅速、准确及复杂的处理。而心理实验的发展,受到以下三方面因素的制约或推动:(1)在对心理学研究的对象和基本任务的理解上发生着变化,如前述各种心理学派对心理学对象与任务的理解就很不相同,所以他们开展实验研究的方法也不相同;(2)实验研究方法的应用范围逐渐推广,无论是在研究的现象和问题方面,还是在研究对象方面都是如此,这种推广对实验心理学提出了更多要求,也因此刺激了它的发展与变化;(3)在实验研究中力图得到精确可靠的和客观的结果,因此实验的设计和控制手段在不断得到改善。显然,随着生命科学的繁荣和现代科技成果的逐渐被采用,21世纪心理学的普及和繁荣是不容怀疑的。

三、实验心理学的逻辑

(一)什么叫做实验心理学

实验心理学(experimental psychology)有广义和狭义之分。在谈到心理学

① 杨治良:《基础实验心理学》,甘肃人民出版社1988年版,第4—6页。

发展的时候,实验心理学作广义理解,它是科学心理学的代名词,是当心理学从哲学的附庸地位独立出来以后,采用实验作为其主要研究方法的心理学体系。它代表了心理学发展的主流,因此从构造主义心理学开始,然后是机能主义心理学、行为主义心理学、完形主义心理学,一直到现代认知心理学,都属于实验心理学的范畴。这种意义上的实验心理学与精神分析、人本主义心理学具有某种不相容性,也即存在着文化分裂。

在实证主义者看来,科学要尽可能贴近可观察的事实及以之为前提的归纳,具有"描述"(description)和"控制"(control)的功能。控制是最终的理性,也是实用主义者所希冀的。因此,科学主义的心理学接受自然科学的发展模式及研究方法,寻求和建立心理活动的一般法则。科学心理学的缔造者之一费希纳借用数学和物理的方法为其"范式论"寻找证据,建立了心理物理学,为科学心理学划定了一个基本范式(paradigm)。冯特明显地承继了化学和生理学的传统,采用分解、还原的方法建立实验心理学体系。华生等人干脆抛开意识,把人看成是"动物"或"机器",刺激与反应的联结一时成为心理学的通用术语。这一研究在追求对心理法则的认识上、对心理和行为的预测及控制上取得了丰硕成果,但同时确实忽略或放弃了对人的情感、动机的研究,忽略了对整体的、有丰富文化内涵的"人格"的研究,这就必然引起人文主义的批判。

实验心理学的初衷在于建立完全实验科学性质的心理学体系,但是迄今这一意愿在某种程度上破产了,实验心理学成为心理学科大家族中的一部分,甚至在许多心理学家的眼中,它就是一门工具性的方法类课程。这也难怪,人不是纯粹自然的存在,纯粹自然科学的方法是无法完成探索人的心灵这一最复杂的存在的运动机制的;人还是社会性的存在,文化的作用机制在这里无处不有,其中的绝大部分成分隐藏于我们感官所能触及的范围之外。"感官"①不及的地方,恰恰是实验方法的沼泽地,人文主义的研究取向理所当然地分享这一神奇的领地。多方法共存是人类心灵探索历史最主要的特征,也是当代多元文化社会中心理学家必须承认的历史和现实。就心理学家个人来说,他必须培养关于心理学研究方法论的最合理的信念,然后选择和思考适合自己的研究取向。

广义的实验心理学是相对于人文取向的心理学体系,也叫做科学心理学。编著本书的目的主要在于分析心理实验的基本程序及其一般的应用问题,故对

① 此处的感官不仅指我们的生理感官,也包括利用各种仪器设备延伸了的"感官"。

实验心理学持一种狭义的理解:实验心理学是研究心理实验的基本理论、基本技术并概括和介绍心理学各分支领域中实验研究成果的科学。具体来说,其内容包括两大部分:第一部分是心理实验的一般理论和一般方法,主要是心理实验的设计、实施,以及伴随这一过程出现的一般问题,比如如何选择研究课题、如何分析和处理研究变量、如何抽取研究对象、如何分析研究资料和呈现研究结果、研究者应遵循的伦理规范等等;第二部分是在具体的心理学分支研究背景下,实验方法的应用问题,它一方面为第一部分内容提供了例证,另一方面将实验方法具体化并因之有利于阅读者研究技能的形成,这里可以包括感知实验、记忆实验、思维实验、情绪实验等等,也可以包括基础心理学的实验研究、发展心理学的实验研究、教育心理学的实验研究、社会心理学的实验研究、临床心理学的实验研究等等。

本书采用对实验心理学的狭义理解,并兼顾心理实验的基本原理和心理实验方法应用两个方面,尽量将二者融合为一个整体,因此名曰"应用实验心理学"(applied experimental psychology)。

(二)心理实验及其逻辑

1. 观察技术与相关技术

观察(observation)、相关(correlation)和实验(experiment)是实验心理学的三种基本技术。其中实验是核心,因为只有实验才能更有效地探明心理活动的规律,探明了规律,才能深入认识人的心理生活并指导人们的心理生活,但是实验技术需要观察和相关技术的支持。

观察法就是对研究对象进行有计划、有步骤、较持久的观察,它是出于描述的目的,而且只能达到描述的目的。观察是一种外部知觉活动,它为了保证得到客观真实的资料,要做到对研究对象不施加任何干预或尽量不带来任何影响,一般采用的方式包括三种:自然观察、参与式观察、结构性观察。自然观察,是隐蔽的、远距离的对研究对象进行观察,保证观察到的活动是自然发生的;参与式观察,就是当隐蔽观察存在困难时,观察者作为活动成员参与到研究对象的活动中去,但要隐蔽自己的身份;结构性观察,也叫做策划性观察,就是创设一定的情境,将研究对象的活动安排在这一情境中。这里有一点非常重要,那就是:被观察的对象并未觉察到活动现场的人为性。

观察技术便于客观地描述研究对象,但是往往较为耗时,而且因为未对任何因素施加控制,也不能探明现象间的因果关系。为了保证观察资料的真实可信,

观察要系统、全面。

相关技术是在观察基础上的一种关系描述，它通过统计技术将变量间的关系量化，测量出变量间关系的密切程度和共变关系的方向。这比观察法进了一步，而且常常加进一些测量方法。但是，相关技术也未对研究对象及其环境施加控制，所以也不能获得变量间的因果关系。不过，相关技术能够帮助研究者发现可能的因果关系，可以为实验研究提供先期信息。

2. 实验技术

实验法是近现代科学，特别是近现代自然科学普遍采用的方法。在心理学中，它就是有目的地严格控制或创设一定的条件引起或改变被试的某种心理现象从而进行研究的方法。与其他的研究方法相比，它有以下三个重要特点：(1)研究者处于主动地位，可以主动引起或改变被试的心理活动，不必消极等待某种心理活动的自然发生；(2)研究者可以控制一切偶发的额外变量，保证研究课题的顺利进行及其有效性，科学地揭示心理活动的规律；(3)可以主动改变各种条件，多次进行重复实验，对实验结果进行核实和验证，保证实验的可信度。

心理实验有两种形式：实验室实验法和自然实验法。实验室实验法是指在特定的实验室里进行的、借助各种专门的仪器设备研究心理活动的一种方法。研究者进行心理实验时：(1)要选择一定的对象，并通过标准化的指示语向这些研究对象宣布他们参加实验的做法，要求其严格按照要求参加实验；(2)研究者必须严格控制其他因素的干扰，使实验条件保持稳定，然后操纵或改变实验课题所选定的一个或多个实验变量；(3)充分利用现代化仪器设备，记录通过实验获得的各项数据，主要是研究对象的心理或行为活动测量数据；(4)分析整理实验资料，对获取的数据进行统计处理和分析，找出其中的因果关系，得出结论。

实验室实验法的局限性主要有两点：(1)对复杂心理活动的研究还有一定困难；(2)由于实验大都是在人为的特殊条件下进行的，常会引起研究对象的紧张情绪和期望心理状态，而这些困难难以控制，这样，实验室中观察到的心理活动在一定程度上失去常态，自然就影响到实验结果的可靠性和实用性。

自然实验法是在个体自然的生活、学习或工作条件下，由研究者主动改变或创设某种情境条件，有意引起研究对象某种心理活动的发生，从而研究心理现象的一种方法。这种方法的优点是：既主动创设条件，又是在自然情境中进行的，它克服了实验室实验法明显的人为性。这种方法兼有观察法与实验室实验法的某些优点。

上述两种方法的相同之处在于，都是在控制某些变量的情况下，通过观察研究对象的反应来探求变量间的函数关系，推断人的心理活动规律，这是心理实验法的本质所在。两种实验法的区别是相对的，特别是在现代心理学研究中，常用模拟性实验的方法进行实验室实验，缩小了实验室情境与生活情境的差异，克服其局限性使研究结果更有推广价值。总之，心理实验法与其他心理学研究方法相比是最严密、最有效和最可信的方法，是我们学习和研究心理学问题不可缺少的一部分。当然，对其局限性我们也要认识清楚，它必须和其他的心理学研究方法相互结合和补充。

3. **心理实验的逻辑**

心理实验是一种科学的方法，科学的方法应该能够克服研究者的主观性，保证研究结论的准确可靠，那么心理实验的科学性体现在什么地方呢？为说明这一点，我们先抽取出心理实验的最一般程序或规则：实验组与控制组实验。

前文说过，人的心理现象不是物理现象，它受到许许多多主客观因素的交互作用，而在这些因素中存在许多随机变化因素，这使得心理现象的变化也具有随机性，所以我们必须将心理事件作为概率事件来研究。这就是为什么在许多情况下，我们不使用单个人作为研究对象，而是使用成组被试作为研究对象。我们研究的是心理现象变化的概率性，为得到因果关系，我们常常使用两个组或更多个组。心理实验的程序可以表示成如图 0-1 所示的通用形式。

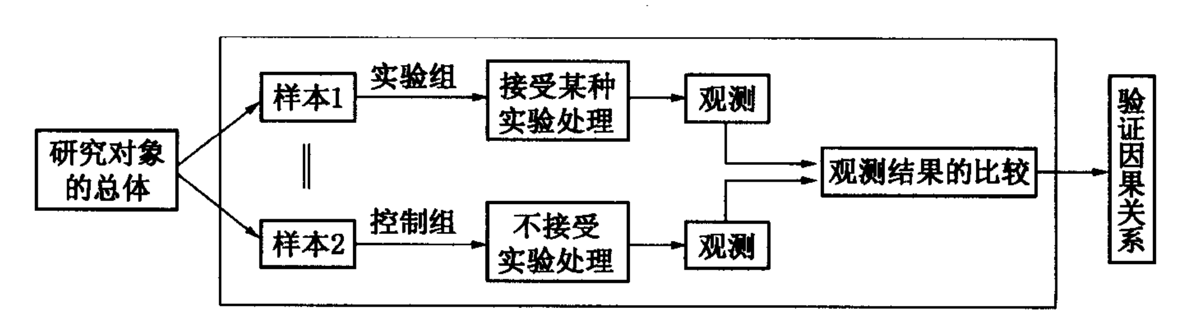


图 0-1 心理实验的通用程序或逻辑

图 0-1 中，研究者从研究对象的总体中抽取两个相等的样本。在实验之前，这两个样本的心理活动水平是相同的，其中样本 1 充当实验组，样本 2 充当控制组。在实验过程中，实验组接受某种实验处理，而与此同时，控制组未接受实验处理，随后对这两个样本中个体的心理活动水平进行观测并比较。因为，实验是在一种隔离的情境中进行的，也就是将其他可能存在的环境变化因子控制在不变的水平，那么如果在两个样本的观测结果中出现了差异，其原因是唯一的，即

是否接受了实验处理，由此可以证实因果关系。这就是心理实验的通用程序或逻辑。

实验方法的应用，使心理学的发展大为加快。虽然科学心理学不能就归为实验心理学，实验方法也不是心理学研究的唯一方法。但任何当代的心理学教科书，都以大量的篇幅证明，现代心理学的大量事实来自实验研究。正如我国著名心理学家杨治良先生指出的："一位心理学工作者可以对心理学的任一领域任一分支特别感兴趣，可以专门从事儿童心理、教育心理、医学心理，或知觉心理、思维心理，以至社会心理的研究，但是如果他想成为一个真正严格的科学的心理学工作者，他就必定要很好地掌握实验心理学的研究内容和方法，了解应当如何科学地考察心理和行为的规律。"①实验心理学是运用实验方法研究心理现象的科学，它概括和总结了心理实验研究的方法和成果，是心理学体系中最基本的组成部分，它代表着心理学发展的主线和发展水平。实验心理学的建立标志着科学心理学的开始，实验心理学发展的三个时期标志着心理学发展的三个阶段。实验心理学既是心理学的基础学科，又是一门心理技术学，它可以在心理学的理论研究和应用研究之间架起一座桥梁。

实验心理学成为心理学工作者的必修课，因为它是心理学家享用广泛领域中心理学研究成果的基本准备。学习实验心理学的目的和意义，就是要得到一些心理学研究方法的训练，掌握心理实验的基本理论、基本技术，特别是掌握心理实验的设计及操作，逐步形成基本的心理实验技能，激发探索激情，培养实事求是的科学态度，并进一步加深对心理学基本概念和基本理论的理解。

阅读材料 0-1

中国大陆实验心理学的发展历程

中国的实验心理学开始于 1917 年，著名心理学家陈大齐教授在北京大学哲学系建立了中国第一个心理实验室。随后有七八个高等学校设立了心理系，开

① 杨治良：《基础实验心理学》，甘肃人民出版社 1988 年版，第 7 页。

设了实验心理学课程，建立了心理实验室，但建国前真正开展心理实验研究的并不多。特别是在抗日战争期间，心理学的教学和科研工作基本处于停滞状态。

中华人民共和国成立后，实验心理学的发展大致可划分为以下几个时期：1951—1957年主要是学习传统的实验心理学，研究简单的感知问题，如中国科学院心理研究所研究了一些运动知觉方面的课题。这时受苏联心理学的影响较大，特别是巴甫洛夫学说的影响。1957—1961年主要进行了一些劳动心理学的研究，这与1957年开展的全民整风运动中理论联系实际的讨论有关。这些研究如炼钢火焰的视觉判断、飞行人员的飞行选拔、飞行错觉、听力零级曲线等等，都有重要的实际意义，但同时也削弱了基础研究，与国际心理学的距离拉大。1962—1966年，我国实验心理学得到较大发展，在大小距离知觉、深度知觉、空间定位、图形知觉、听知觉、触觉、动觉、感觉相互作用及错觉方面，都有比较系统的高水平的论文发表，研究范围扩大了。从1966年的"陈葛辩论"开始，即1967—1976年，心理学的研究工作受到严重破坏，只有少数心理学家在悄悄地进行一些研究，航空心理学是"文革"期间唯一未被破坏的心理学分支。"文革"以后，实验心理学很快得到恢复，1979年正值心理学诞生100周年之际，中国心理学会实验心理学专业委员会成立，中国科学院心理研究所和高等学校的心理系、教育系也都开始了比较系统的心理学教学与研究工作。

从80年代中期以后，逐渐与国际心理学的研究潮流接轨，并出现几种重要趋向：(1)接受并开展认知心理学的研究成为主流；(2)心理学实验仪器的研制出现规模；(3)各心理学科点开始加大投入，建立心理学实验室，并加紧了硕士生、博士生的培养；(4)国家自然科学基金的资助使心理学研究出现专题化，形成了几个内容不同、风格不同的实验心理学研究群体。这些变化给中国心理学带来巨大发展，许多有价值的研究成果得到国际心理学界同行的赞赏，1992年中国心理学家荆其诚先生当选为国际心理学联合会副主席。

特别是，2004年8月8日至13日，第28届国际心理学大会在北京召开，这是中国心理学走向世界，中国心理学家的工作越来越得到国际心理学界承认的重要标志。此次会议在北京亚运村国际会议中心召开，会议规模空前。会议注册人数达到6 000多人，其中国外代表3 500人，国内代表近3 000人，另外还有许多未注册的研究生到会听取会议发言。此次会议得到国际心理学界的高度重视，诺贝尔奖获得者卡尼曼教授到会并在开幕式上作了学术报告。中国科学院院长路甬祥出席会议并致词。与此同时，国际心理学联合会在北京召开了工作会议，在此会

议上，中国心理学会理事长张侃研究员当选国际心理学联合会执行委员。

当前我国实验心理学研究的课题涉及心理学的各分支领域，在人工智能、记忆、思维、元认知、神经生理、认知神经科学等国际心理学的前沿领域都取得了很多成果。

——主要参考资料：《中国心理科学》（王甦等，1997）①

阅读材料 0-2

中国台湾实验心理学的奠基者：郑发育

郑发育教授1916年生于台南市。18岁赴日留学，就读高等学校、大学及研究所。1946年自京都帝国大学大学院的哲学研究所返台，旋即应聘于台湾大学哲学系。1949年该校创设心理学系，他改隶该系。1953年升为副教授，1956年到美国耶鲁大学米勒（Miller, N. E.）教授的实验室研究一年，1959年升为教授，1986年退休。1996年患脑溢血去世。心理学系创系初期，资源短缺，师资、财源均甚拮据，薪水亦低。在此困难时期，他安贫乐道地尽心设法让实验心理学的课能够开成，使学生得以体验用实验的方法来研究心理学；并且训练陈台先生成为当年台湾唯一的心理学仪器的制造者，使早期的台湾心理学界得以进行实验研究。此外，在那个年代，他经常到系研究，不但是学生最常请益与模仿的对象，而且也是发表研究论文最多的教师。由于他的身教，使实验研究成为心理学系的传统，将台湾心理学的研究奠基于科学方法。在台湾，他不但是实验心理学研究的祖师，而且也是第一位用实验方法来研究人格心理学与临床心理学课题的心理学家。

台湾大学心理学系创立之初，师资极端缺乏，郑先生先后开过知觉心理学、社会心理学、人格心理学、学习心理学等必修课程。但是，最主要的是实验心理学。郑先生除了规划和教授各个实验的主题外，还亲自从日据时代遗留下来的150件实验仪器中挑选与安装，供学生实际操作，并要求学生依据所得的结果仿照学术论文格式撰写报告。因此，学生必须阅读相关的学术文献，得以实际体验到用科

① 王甦等：《中国心理科学》，吉林教育出版社1997年版，第1—18页。

学的方法来探索心理学的课题;不但如此,他还要求学生上台报告,鼓励同学批判,体验科学的严谨要求,因而训练出批判的精神。这种科学与批判的精神遂成台湾大学心理学系的传统,并经过该系毕业生的传播,影响到台湾其他后来设立的心理学系,也就是说,影响到台湾的心理学界。

最后一件是技术人才的培训。实验心理学方面的研究工作不能只靠日据时代留下来的仪器。特定的研究课题需要特定的仪器。在当年财政极端困难的情况下,新购仪器几乎不可能,自制是唯一的办法。1951年左右,台北工专电机科毕业的陈台先生到台湾大学心理学系任技工之职,负制造仪器之责。郑先生乃设计仪器,指导陈先生制造各式各样的仪器,将陈先生训练成台湾唯一的心理学仪器制造者,集电工、木工及金工的技艺于一身。陈先生的能力,不但支持了实验心理学的教学需求,而且也帮助研究生能顺利完成学位论文的实验,更是刘英茂先生与其他系内系外心理学教师进行研究工作的重要靠山。台湾实验心理学的研究工作能够奠基,陈先生功不可没,而"幕后英雄"就是郑发育先生。

——资料来源:《台湾实验心理学的奠基者:郑发育教授》(徐嘉宏等)①

建议阅读文献

1. 高觉敷:《实验心理学的诞生和冯特的贡献》,见高觉敷主编:《西方近代心理学史》,人民教育出版社1982年版,第115—137页。
2. 叶浩生主编:《西方心理学的历史与体系》,人民教育出版社1998年版,第54—91页。
3. 《中华人民共和国成立40年来的心理学发展》,见王甦等主编:《中国心理科学》,吉林教育出版社1997年版,第1—18页。
4. 张学民、舒华编著:《实验心理学纲要》,北京师范大学出版社2004年版,第1—12页。

复习思考题

1. 如何理解实验心理学、心理实验、实验室实验法、自然实验法、自我观察法(内

① 徐嘉宏、吴英璋、余德慧:《台湾实验心理学的奠基者:郑发育教授》,《中华心理学刊》,41(2)。

省法)?

2. 科学方法的主要特征有哪些?
3. 学习实验心理学的目的和意义是什么?
4. 简单介绍实验心理学的发展史。
5. 简单评介费希纳对实验心理学的贡献。
6. 简单评介冯特对实验心理学的贡献及局限性。
7. 影响和制约心理实验方法发展变化的因素主要有哪些?
8. 中国实验心理学的发展经历了怎样的过程?
9. 你是如何理解心理学的理论研究与实证研究的关系的?

第一章
心理学实验研究的过程

本章内容提要

心理学是一门实验科学，要理解心理学，就要理解其实验研究的方法。掌握实验研究的程序，不仅是开展心理学研究的需要，也是阅读、评估、分享他人成果的需要，甚至还是在生活历程中作出各种正确决策的需要。本章内容主要包括：(1)课题选择，讨论研究课题的来源：文献检索与分析、个人兴趣和经验、实际需要和实践中的矛盾、基金项目申报的课题指南等；(2)实验设计，讨论实验中变量的分析与处理、被试的选取、实验设计模式的选择等；(3)实验程序的实施，主要讨论实验实施过程中主试与被试的交互作用，以及这种作用对实验研究结果的可能影响；(4)数据的处理与分析，简单分析实验结果的描述性统计分析和推断性统计分析的功用；(5)研究报告的撰写，介绍研究报告各部分应表达的内容、要求与技巧。

在开始讨论心理学实验研究的一般过程之前，要先解决学习者往往会存在的一些困惑：我可以不学习实验心理学吗？毕竟我想成为一名心理学家而不是科学家，或者至少我不想成为从事实验研究的心理学家，那么，我是不是就可以不学习心理学实验研究的程序了呢？不少学生认为实验心理学既枯燥又难学，还是绕道而行的好。但杨治良先生告诉我们，实验心理学具有基础价值，[①]绕道而行并非上策，在将来的职业道路上你会经常与之遭遇。中外学习心理学的学生起初大都具有这种心态，这种心态可能源于对科学的敬畏。这种敬畏的心态首先会阻止学习兴趣的发展；当然，在遇到困难时，它又成了退却的借口；在成绩不理想时，它还是保证能心安理得的防御体系。

---

① 杨治良：《基础实验心理学》，甘肃人民出版社1988年版，第7页。

我想说的是：心理学实验研究的过程是如此简单明了、有趣，认识它，你才会喜欢它；喜欢它，它才会给你带来欢乐！

第一节　了解实验研究过程是必要的吗

心理学专业的学生一开始不喜欢方法类课程，尤其不喜欢其中的统计学，这是完全正常和可以理解的。他们选择心理学专业是因为想了解人，而研究方法不是关于人的科学，它当然也不是心理学，它是科学。同样，统计学也不是心理学，它是科学家们使用的数学。① 令人懊恼的是，实验心理学属于方法类课程，还偏偏包含大量统计学！

现在来思考以下几个方面的问题，这会使你明白：了解心理实验研究的过程肯定会是有用的。

一、心理学是一门实验科学

为什么要学习实验心理学呢？因为心理学首先是一门实验科学。心理学家在解答心理学的许多问题时要借助实验方法，于是他必须按照某一确定的程序来收集资料、分析资料和得到答案。假如一位心理学家想研究酒精和咖啡对人的反应时间有无影响，他就需要对喝过酒的人和喝过咖啡的人进行反应时间的测量。但是，如此测量还不够，因为你不知道测量得到的反应时间是在喝酒或喝咖啡时才有的，还是平常就如此。于是，你需要找来另外一些人参加反应时间的测量，这些人在测量前某一定时间内既没有喝酒也没有喝咖啡。然后对这两部分人的反应时间进行比较，看是否存在差异。如果存在差异，那么是否就可以得到结论呢？这还要看他们之间除是否喝酒、喝咖啡之外还有没有其他的不同。这样的程序是繁琐了些，但它可以保证心理学家查明喝酒和喝咖啡与反应时间之间是否真的存在因果关系。

今天心理学教科书中的许多知识都是依靠这些程序积累起来的，这种程序就是心理实验。实验心理学为解答心理学的疑难问题提供了开放的和完备的程

① ［美］弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著，邓铸、姜子云、蒋小慧等译：《行为科学研究方法》，陕西师范大学出版社 2005 年版，第 2 页。

式,它可以保证得到尽可能准确和完整的答案。想成为心理学家需要这种科学方法,只是想理解心理学和完整地享用心理学家关于人的发现,也需要对实验研究方法有一个基本把握。

二、开展实际的研究

选择了心理学专业,或者心理学的相关专业,你很可能会在某个时候,也要开展一项实际的研究。比如,在你后续的许多课程中,主讲老师会要求你独立地完成一些研究,或者至少在你将要大学毕业时,你需要提交一份研究报告,这时你就需要心理学的实验方法或类似的其他方法了。此外,你在毕业之后,或许会得到一份与人的研究有关的工作;或者,毕业前,你就想进行一些研究,以增加你学位的分量,这些时候也需要心理学的研究方法了。现在,有不少学生想在大学毕业之后继续攻读硕士学位或博士学位,那时心理实验方法就肯定是要学会的了。而如果你想在大学期间增加考取研究生的竞争筹码,你可以进行一项独立的研究。

当然,许多学生会觉得“心理学的实验研究”离他们还远着呢,因此认为实验研究方法课程与满足他们的近期需要无关。而且,有些学生并不打算以后从事研究工作,他们只想在大学毕业之后能在服务行业得到一份好工作。因此,不少学生认为实验研究方法与他们的职业兴趣也毫无关联。即便如此,实验研究方法课程还是有用的,为了不在工作中落伍,你需要经常阅读和理解最新发表的研究论文。可是,如果没有关于研究过程的知识,要想读懂这些研究报告很难。

三、阅读和评估他人的实验报告

掌握实验研究的术语和逻辑使你能够阅读和理解实验研究论文,这不只是说你要能够阅读杂志、报纸或教科书上别人所作的关于其他人研究的总结,而是你能够亲自阅读原创论文并从中总结出你自己的结论。研究方法课程还将帮助你阅读并批判性地评价杂志上那些详细记载研究过程的论文。许多职业都要应用研究的成果,比如说你是一位临床心理学家,试图确定在临床上使用哪一种治疗方法最好,你可能就会翻阅有关研究论文以考察不同治疗方法的效果。类似

地,如果你是一位教师,试图确定最适合学生的教学方法,你也可能会翻阅研究论文以考察不同教学方法的效果。阅读和评估研究论文有助于你确定最适合临床病人症状的治疗方法或最适合你的学生的教学方法。此外,其他课程的学习也需要经常性地阅读原创性研究报告。

理解研究术语也有助于你严格地评估杂志上发表的论文。许多论文都会从"结果"部分(论文的此部分是告诉读者研究中发现了什么)过渡到"讨论"部分(论文的此部分是作者对结果进行解释并得到结论),这种过渡非常必要,它有助于发掘结果中隐藏的"理论",但这种过渡也容易犯主观错误,它容易夸大研究的价值,或将读者引向错误的结论。所以,你必须能够分析和评判这种过渡。研究方法课程可以帮助你评估他人的研究报告。

四、理解实验研究简介

实验研究方法课程也将帮助你理解简单的研究介绍。在大部分的心理学课程和教材中,都会提供一些研究的简单描述来作为支持某些结论或理论的证据。比如,向你讲述一项研究:以治疗方式作为自变量、吸烟量作为因变量,采用安慰剂控制组的组间设计进行实验,即一组烟瘾者服用烟碱片(nicotine patch),而另一组烟瘾者服用不含烟碱的安慰剂药片。研究者发现烟碱片可以显著地减少重度吸烟成瘾者的吸烟量。正如你在这个例子中所看到的,一本教材或一位教授在向你描述别人的研究时,不会将有关这个研究的每个细节都告诉你,而是使用一种心理学家用来描述研究的固定模式(行话或术语),这种模式就是由研究方法学的基本原理确定的。

你在这本书中学习到的一些基本原理是常用的、基本的,是许多研究都可以遵循的。因为许多研究都遵循这些原理,所以作者会假定阅读者也熟悉这些原理,在描述他人的研究时就比较简洁。方法课程可以帮助你填补关于某项研究的描述中通常留下的空白。可是,如果你恰恰没有系统学习过这些方法学,也就不理解这些研究方法,你就会觉得有些实验简直莫名其妙。例如,上面描述的这个实验中,为什么必须让一个烟瘾者被试组(安慰剂控制组)服用不含烟碱的药片?研究方法课程可以帮助你更好地理解和记住一些研究,随后你就能更好地掌握课程中的其他内容。

另外,即使你准备从事理论研究,可能也需要掌握实验方法,它可以帮助你

对大量文献作出自己的评判和选择,尽量避免人云亦云。

五、日常生活中的决策

理解研究过程具有广泛价值,而不只是对研究者或心理学专业人士有用。要成为21世纪一个有能力的、有积极参与意识的社会成员,就必须能理解和欣赏科学和实验的力量,以便能够对研究结果进行评价并依据研究成果决定自己的行动。

假如,今天报纸上刊登了一则介绍催眠减肥程序的广告。为证实这一程序的有效性,这则广告介绍了三名满意的参加者:杨大嫂,她改变了饮食习惯,体重下降18公斤;蔡大妈,体重减少20公斤;叶小姐,从刊出的其在治疗前后的照片可以明显看出,四个月减掉13公斤的效果。这科学吗?这些对减肥程序的检验可信吗?答案可能是肯定的,也可能是否定的。如果杨大嫂、蔡大妈和叶小姐真的是这一训练程序使用的临床患者的代表,那么她们的结果是在正确的方向上迈出的一步。但如果情况不是这样,减肥指导医生有意选择的是那些临床上特别成功的案例,那么这肯定就不是科学的研究了。或者假定,我们从临床记录中找到一些案例,她们在采用这一减肥程序后出现了病症或意外事故:小王在刚开始采用这一减肥程序两个月后从梯子上摔下来造成腿骨骨折;小马在接受6个星期的催眠训练后患了急性阑尾炎,住院5天……读到三四个此类的故事后,你是不是就确信这种减肥程序对人的健康是有害的呢?情况也未必如此。

每天,来自报纸、杂志、电视和无线电广播中的各种研究声明像洪水般将我们淹没,什么"研究者已经发现……","研究显示……",什么"油炸食品会提高患癌症的风险","服用海胆亚目(echinacea)可以减轻感冒症状","每天喝一杯葡萄酒可以降低患心脏病的危险","食用含高碳水化合物的食物可以降低体重"。

我们该如何面对这些信息呢?是加以接受而改变我们的行为——停止食用油炸食品、开始服用海胆亚目、每天喝一杯葡萄酒和吃一些面包或其他面制食品,还是置之不理?作出哪种选择是最好的呢?我们要想更好地享用这些研究声明,那就需要接受研究方法的训练,这有助于我们找到并评估那些据称是支持某一声明的原创研究(如果确实有这些证据)。一个外行也许可以凭借批判性的和符合逻辑的思维去发现某些数据收集或逻辑上的缺陷。研究方法课程可以让你知道在进行研究和数据解释中的逻辑约束,以便你能够依靠自己从数据中梳

理出真实的结论，而不再仅仅依靠所谓的专家为你得到的结论。研究方法课程可以使你针对每天生活中的各种观点作出理性的选择。

最后，我们说科学方法不仅是一种研究方法，是一种能够获取问题答案的方法，而且也是一种符合逻辑的、理性的批判性思维方法。这种思维方法不只是针对心理学而是可以应用于生活中的所有方面。研究方法课程将教会你像科学家那样进行思维，而且我们希望你能明白，不是只有实验研究者才需要这样的思维。

第二节　实验研究的起点：课题选择

心理学的实验研究是一项结构严密、步骤分明的工作。具体地说，可将其划分为五个主要阶段：课题选择、研究设计、实验实施、结果及其分析、研究报告的撰写等。此外，在研究工作的最后，还要对整个研究过程进行总结与反思，这或许又是下一研究的起点。实验研究的每一阶段都有技术性很强的工作要做，而且由于其中几乎每一阶段都涉及人的因素，所以又常常要求其符合伦理考量。本章以研究的五阶段为路线，系统讨论每一阶段的技术和伦理问题。

讨论的起点自然是课题选择(subject selection)，这是研究工作的基本单元与起动，也是研究者团体或个人的经常性工作。能否选定一个合适的研究课题，对于研究工作能否顺利展开和能否获得满意的成果具有重要意义。如果研究主题不明确或选题不当，研究工作就会混乱、盲从，事倍功半，甚至白白耗费人力、物力和时间。课题选择还规定着研究需要采用的方法，因为某一确定的课题，必须采用与之相应的方法和手段来研究。对于内隐的心理机制，不能采用像利用解剖方法直接观测人脑内部的生理结构那样的手段来研究，而是需要采用像计算机模拟或类似的“黑箱”方法来间接地进行研究，并借助理论思维对其做出推断。由于选题在研究工作中的特殊地位和作用，每一位研究者都非常重视。一般来说，提出课题比解决课题更困难。

课题来自何处，如何选题，这是首先要讨论的问题，为此我们对下列课题来源进行分析。

一、文献的检索与分析

科学研究是在继承基础上的创造性活动，撇开前人丰富的研究文献而盲目

单干肯定是不明智的。文献检索(literature search)与分析是提出有价值的前沿课题的基础性工作,也是研究课题具体化的必经之路。文献检索与分析存在两个层次:第一个层次在于把握研究生涯的方向;第二个层次在于确定具体的研究课题。

第一层次的文献检索与分析要面向整个心理学学科和某一分支学科发展格局,这往往是进行大型选题和对自己一个时期甚至整个研究生涯进行定位时必须花费较多精力去做的事情,目的在于考察历史、反思过去和展望未来。这一过程要求研究者对整个学科发展现状、方法论体系和未来趋势形成自己独特的看法。为使这一过程更为有效,需要借鉴或采用下述文献检索与分析的策略和方法。

(一) 注视学科发展的生长点

密切注视现代心理学发展的趋势,结合国情,不失时机地抓住学科生长点上的新问题,是选择新课题的重要策略与思想方法。前沿性课题对学科发展具有战略意义,当然也常常存在激烈竞争,同时会有几个、几十个,甚至更多的研究者进行竞争性研究。与其他研究者同时挤进一个狭窄的领地去开垦,如知识储备不足、判断不准、决策不优,都会在竞争中失败。因此,研究这类前沿性问题具有很大风险,要想取得成功,必须有充分的知识储备和接受过严格的科学训练。

(二) 寻找学科研究中的空白区

科学研究中能得到最大收获的领域是长期被人忽视的空白区(无人区),那里因长期无人涉足,大多数研究者不敢轻易闯入。然而,无人区里任纵横,存在着大量的机遇。勇敢的闯入者能在这无人开垦的领地中搜索到值得研究的问题。诚然,到科学的“无人区”去寻找问题、发掘新课题,必须对当前的学科知识背景和个人的主观条件作出客观估量。那些超前于学科发展水平且不具备条件的新课题,或者超出个人学术能力的课题,不宜贸然选定。因为这类课题凭借现有的主客观条件很难在可预见的未来获得像样的成果。

(三) 注意心理学与临近学科间的“结合部”

早在19世纪70年代,恩格斯就以深邃的眼光,洞察了物理学与化学的内在联系,他在《自然辩证法》里曾告诫研究者要注意不同学科之间的“边缘区”:“在分子科学和原子科学的接触点上,物理学家和化学家都承认自己没有能力,然而就正应在这点上可望取得最大的成果。”① 由于现代心理学是一门交叉科学,它

① 恩格斯:《自然辩证法》,人民出版社1971年版。

与许多自然学科和社会学科之间存在接合点,因而有着许多充满生机的“边缘区”和“结合部”。只要研究者密切注视这些“边缘区”和“结合部”,值得研究的问题就会层出不穷。这一点不难理解,当代认知心理学与神经生理学的有机结合,诞生了当代心理学领域中最具魅力的研究领域:认知神经科学(cognitive neuroscience)。毫无疑问,认知神经科学也成为当代心理学研究中成果最为丰富的领域之一,她成就了一大批中外心理学家。

(四) 运用其他学科的理论与方法审视心理学问题

科学问题的解决总是相对的、不彻底的,在一定的科学知识背景下被认为已经得到解决的问题,在另一种知识背景下,常常又会重新成为问题,需要从新的角度进行研究和解决。尤其是那些长期悬而未决的难题更应利用现代科学所提供的背景知识,加以重新审视。因为沿着一条思路往前走,固然会有所造诣,但终究会遇到一些无法超越的障碍。这时如能重新打开一条或几条新思路,运用其他学科的理论和方法,审视长期无法解决的心理学问题,也许能在诸多研究路线的交汇点上,寻找到有利于问题解决的新途径与新方向,正所谓“他山之石,可以攻玉”!

(五) 揭示已有心理学资料间的内在联系

近年来,由于观察、实验研究(也包括大量的测量研究)积累起浩繁的经验性资料,心理学各分支学科都面临着资料整理、概括和分析的艰巨任务,即从经验材料出发,通过理论思维,揭示其中的本质联系(内在规律)。如果心理学的实证研究滞留在观察、描述和数据记录阶段,不去发掘资料间的内在联系,探索支配心理活动现象的内在规律,那么实证资料就会失去它存在的应有价值。所以,充分利用心理学中长期积累的资料,对其进行整理、概括和分析,是心理学研究的一个重要方面,也是心理学研究课题的主要来源之一。

需要强调的是,心理学既需要理论思维,也需要实证技术。心理学研究需要各种力量的相互沟通和接纳,因为她非常需要借助理论思维来整合资料。

(六) 接受哲学预见和科学幻想的启迪

以先进的哲学思想和古老而经久不衰的哲学原理作为心理学研究的哲学背景,或重视科学幻想的启迪作用,有时对于确立新思想、提出新问题是非常有益的。如格式塔心理学关于“整体和部分”的研究与古希腊哲学家亚里士多德(Aristotle,前 384—前 322)的“整体大于部分之和”的思想很有相似之处。尤其是,随着科学技术的发展,由哲学预见和科学幻想变为现实的周期越来越短,为此,

研究者更应直接或间接地从中获得教益。

一个心理学工作者在其研究生涯的初期，虽不能过于空泛地涉猎心理学所有分支的有关文献资料，但是对于一些带有理论指导性的文献，特别是一些反思性、预见性理论文献要有所把握。比如就当前来说，世纪之交的心理学，无论是从整个学科，还是某一分支，都在悄悄地或已经公开地发生着重大变革，如不重视研究取向的研究，就可能使研究走向死胡同。拿认知心理学来说，《认知革命的未来》①就是一部很有启迪意义的反思性、预见性著作，是值得认知心理学学习与研究者去读的著作。

不过，初涉研究者也忌泛而不专。如果蜻蜓点水，没有明确的主攻方向，必定会浪费宝贵的时间，使研究在低水平上徘徊。所以，在尚不能判断应选择什么样的研究方向时，一方面可以从导师那里得到建议，另一方面可先选择一个专题作为研究方向，以期从中获得锻炼并逐渐站到高处去审视学科发展。

选择某一学科的某一方向后，就要进行全面而深入的文献研究和理论分析，这是文献研究的第二个层次：分析特定研究方向的理论假设及其依赖的研究资料，或者考察有关理论的解释效度。不同理论假说往往是不同研究者从自己拥有的实证资料中抽象概括出来的，它们之间可能互有继承的关系、可能基本一致、可能互相补充、可能完全对立。这时，对各理论假说做出正确分析与估价，对于找到自己研究的切入点非常重要。对某一专题领域已有的理论进行分析，并期望据此引申出新的研究课题，可以采用以下的策略与方法。

1. 试图寻找多种理论假说之间的差异和矛盾

多种理论假说同时并存是现代心理学的重要特征之一。为此，研究者可以把不同的学术论文或者不同的教科书进行比较，从中发现多种理论假说对同一问题的不同表述和解释，以及这些解释之间的实质性差异。寻找这种差异，特别是其中相互矛盾的地方，有利于启发新思路、发现新问题、引出新课题。

2. 尝试补充和完善现有理论

理论的完善是相对的，任何现有理论都存在着各种可能的缺陷和需要发展的成分，因而思考怎样使现有理论不断得到补充与完善，必然会引出许多有待进

① Johnson, D. M. & Erneling, C. (Eds.)(1997). *The future of the cognitive revolution*. Oxford University Press, Inc.

一步探索的心理学问题。

3. 追求理论的普遍性和逻辑简单性

追求理论的普遍性和逻辑简单性是理论考察中的一贯传统。能否用一种统一的方式解释众多现象？能否使理论本身更为简洁，用尽可能少的基本概念和假设来概括尽可能多的现象？这类思考有利于形成进一步探索的问题。

4. 揭露理论体系内部的逻辑困难

不少心理学理论在体系方面常常存在逻辑困难：或用作推理的前提并未获得充分的佐证，致使整个理论坐落于“沙丘”之上；或存在逻辑跳跃与推理不严密，获得的“逻辑结论”实际上并非真正从前提中导出；或在逻辑上不能自恰，甚至从同一前提出发，导出相互矛盾的命题，从而造成“佯谬”与“悖论”。因此，研究者在面对现有理论时，要善于思考、分析，敢于提出诘难或质疑，试图揭露其中可能存在的种种逻辑困难，并设法加以纠正，这也是发现问题、形成课题的重要途径之一。

5. 尝试解决原有理论与新事实之间的矛盾

获得了新的心理学经验事实，而原有的心理学理论又无法对之加以合理解释，于是原有理论就会面临难题与危机。比如，如何使原理论与新事实取得和谐一致，消除两者的矛盾和冲突，或者怎样变革原理论，使之具有更大的解释效度。此类思考常能引导人们进入研究并获得发现。

文献检索与分析是研究的基础性工作，是直接影响研究选题质量及整个研究水平的关键步骤。研究者要善于收集、选择、整理、消化和概括文献。文献检索与分析的作用是明显的、多方面的：它有利于使研究者全面把握某一主题领域当前的知识状况；研究热点与争论；存在的空白区；重要而尚未得到充分研究的问题；得到研究但结论不一致的问题；可以继承和需要改进的研究范型与操作方法等。

特别提醒一点：在许多已发表的研究报告的“讨论”部分，作者会提出进一步研究的“设想和建议”，这种“设想和建议”也可为你所用。

二、个人的兴趣和经验

从自己的兴趣开始，选择一个真正感兴趣的主题领域，包括感兴趣的现象或矛盾。从兴趣开始“有助于你保持积极性、减少倦怠感，而且更有可能找到你愿

毕生研究的课题”。[①] 那么,如何去发现自己的兴趣点呢？下面的几种考虑角度无疑是有帮助的。

第一,从特定的潜在研究对象的总体或群体开始考虑,确定你最愿意研究的对象群。如人或动物,婴幼儿、学前儿童、学龄儿童,中小学生、大学生、年老者,航天员、飞行员、运动员、采煤工人、服刑人员等。

第二,从特定的心理现象或行为倾向开始考虑,确定你最愿意研究何种心理现象、何种行为倾向或某种特殊症候。如强迫症、考试焦虑、抑郁症,语言发展、思维发展、社会化,多动、诚实、饮食过量、利他行为、颜色偏好等。

第三,从特定的专题开始考虑,确定你感兴趣的研究领域。如心理健康教育、元认知、老龄化、工作压力、儿童虐待、家庭暴力、人格、学习、动机、情绪、人力资源管理、护理心理、人因学等。

从不同角度罗列可能的研究主题,可以帮助你发现自己的兴趣点。当然,有时,你的兴趣是突然“冒出来”的,比如在看期刊论文的时候,看科普读物的时候,看文学作品的时候,在听课的时候,听学术报告的时候,在与同行或不同行的朋友聊天的时候……对于突然“冒出来”的兴趣要进行过滤,然后就其中少数对你来说可能比较重要的部分作较深入的思考。这里的关键在于你确实想更多地去了解一个“专题”,这才是你的兴趣。在感兴趣的领域进行工作,你才会感到非常有趣和富于激情,要不然,你的研究热情可能会很快消退。

此外,从自己的日常生活、学习、工作和社会交往经验中也会找到有意义的研究课题,“处处留心”对于心理学研究者来说更有意义。前文已经说过,人类的好奇心是科学发展的内在动力,对于一个研究者来说,保护自己的“好奇心”,保持开放的“大脑”,会随时从经验中提出重要的研究课题。如:既然人的智力与脑关系密切,为什么有些“小脑袋”的人会很聪明？“法轮功”痴迷者为什么会杀死自己的亲人并为此感到欣慰？聪明的人为什么会在销售“刷钱药水”的骗子面前出奇的愚蠢？有的人在极度悲伤或惊吓状况下为什么会出现“失忆症”？等等。

三、实际需要和实践中的矛盾

满足社会生活与生产实践的实际需要,是心理学研究的出发点和归宿,据此

① [美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著,邓铸、姜子云、蒋小慧等译:《行为科学研究方法》,陕西师范大学出版社 2005 年版,第 24 页。

可以找到大量的研究课题。例如:(1)运用心理学原理,设计相应的生产环境和新型的工业产品。生产过程中,怎样运用已有的心理学原理,设计出有利于提高生产效率的工作环境,或者在产品设计、工程设计中认真考虑人的心理因素,使人和机器达到最佳匹配。这类旨在解决实际问题的课题既可以导致基础研究,也可以导致应用研究,有些研究还可以为社会、为研究者带来经济效益。(2)利用激励机制,调动人的积极性,提高工作效率。对工作效率的影响,除硬件之外,人内在的愿望和动机也起重要作用。为此,怎样利用激励机制,调动人潜在的积极性,以提高工作绩效,实现既定目标,是管理心理学的一个重要研究领域。(3)优化心理素质,保障心理健康。如何优化人们的心理素质,防止心理异常和不良行为的产生,或者如何矫正和治疗患者的心理疾病,使其恢复心理健康。国内这一方面的研究正得到加强。(4)人际交往中的动机、社会意向、价值观念、态度倾向等。人际关系极为复杂、多变,只要研究者密切关注人际交往中反映出的种种倾向,就能收集到大量值得研究的问题,而这类问题的研究对于维持和协调良好的人际关系是非常必要的。(5)提高教学质量和学习效果的研究。(6)认识人的潜力(工作能力、天资、特殊才能等)和发展机理。在人才选拔、就业指导、工作安排、干部考核、工作评价中都需要对人进行各种类型的心理研究和测评。

四、基金项目申报的课题指南

基金项目申报的课题指南是一个选题捷径,这类指南中的课题具有这样一些特点:(1)它来自相关领域中的专家的提议。一般"基金"管理委员会都会定期向有关专家征集选题,然后组织论证、筛选、公布,所以这里的选题在某种程度上代表了学科发展的方向和前沿,具有明显的学术价值。(2)这类课题往往考虑到社会发展对教育、文化建设、经济调控、意识形态、科学技术发展方面的要求,具有明显的应用价值。(3)这类课题中的一部分是开放性的,即它规定了一个大的课题领域和选题范围,研究者可以根据自己的研究条件和知识准备自行开发子课题。这样,在基金资助之下,就会形成一个课题组和一个研究群体。(4)研究者可以定期地申请资助,如能获得批准,则可以得到一定的资金资助,保证研究的顺利进行。

国内,就心理学而言,可以申请的研究基金主要包括国家社会科学研究基金、国家自然科学研究基金、全国教育科学规划研究项目基金、省市自然科学研究基金、省市社会科学研究基金等。此外,还有不同级别的招标课题。

课题来源是多方面的，除上文谈及的，对于学生来说，其课题还可以来自导师的命题。不管是自选课题，还是导师命题，文献检索与分析都少不了。即使在课题选定之后，也需要在文献检索与分析中将课题具体化，甚至将研究课题转换为假设。假设的形式是“如果 A，那么 B”，这样的假设必须是可以被检验的。所谓可被检验，就是有可能被证实为真，因此可以接受；也有可能被证实为假，因此可以拒绝。既不能被证实为真，也不能被证实为假的，就不能作为实验研究的假设。比如，某研究者想研究家庭教养方式与人的攻击行为的关系，他就可以提出假设：“如果人们接受的家庭教养方式不同，其攻击行为的频率与方式就会不同”，这一假设就是可被验证的。诸如“如果一个人干过的坏事越多，他/她进入天堂的可能性就越少”、“如果人会飞，那么人们的抑郁症就会减少”的一些假设就是不可验证的，因此不能成为研究假设。

在课题选择中，除顺应兴趣外，还需要理性，注意评估课题的研究价值和研究的可行性，这可使你的研究生涯更有意义和更有成就。在选择了感兴趣的主题领域，或者有了想探讨的选题之后，还必须对其研究价值及研究可行性进行初步评估，包括对该领域或该课题研究的学术价值和应用价值、是否具备研究的主客观条件进行初步评估。这种评估能快速地剔除那些研究价值不大，甚至根本没有研究价值的课题，也可以帮助排除研究者的实验条件、知识准备根本不能胜任的那些课题，提高选题的有效性。

例如，你对人们的智力水平感兴趣，想研究不同人群智商的差异性。但是这一课题早就被研究过，再进行探讨已无新意，学术价值很低；人们已经普遍发现，正常人群的智商高低不是其取得学习或工作业绩的最主要的决定因素，研究的应用价值也不大，这一课题就可以放弃。你听了某位专家所做的学术报告后，对认知神经科学非常感兴趣，尤其想探明人在顿悟发生时的脑事件相关电位变化，然而，你的认知心理学学得并不扎实，对认知神经科学更是知之甚少。很明显，你的知识准备不充分，无法胜任这一课题。即使你这些知识储备充分，但是无实验设备可用，也是不行的。对知识准备和实验条件都不具备的课题，必须放弃，或者暂时放弃，或者先创造条件。

第三节　实验设计

在课题选定后尚未实施前，要进行实验设计(experimental design)。所谓实

验设计，就是对实验可能涉及的各方面因素进行系统分析、筹划和安排，以保证后续实验的顺利进行和实现实验研究的目的。实验设计要考虑的主要有三个方面的问题：变量(variable)的分析与处理、被试的选择与分配、实验设计模式的选择等。此外，还要给出实验操作的详细步骤、所需仪器设备和实验材料等。

在讨论实验设计之前，这里先交待主试和被试的概念。主试就是主持实验的人，即直接操控和观测实验中的各种变量的人，也叫做实验者(experimenter)。实验者可以是研究者，也可以是研究助手。被试(subject/participant)就是在研究中被观测的人类或非人类实验对象。① 实验对象或被观测的对象不等于“研究对象”，而往往只是研究对象中的一部分。通常，潜在的或可能的研究对象的全体叫做总体(population)，而从总体中抽取出来实际参加实验的人或动物则叫做样本(sample)。进入到样本中的人或动物才叫做被试。

一、实验中变量的分析与处理

(一) 心理实验中的变量及其分类

所谓变量，就是可以在数量上或性质上发生变化的事物的属性。根据其来源，心理实验中的变量可以分为三类：刺激变量(stimulus variable，常以 S 表示)、机体变量(organism variable，常以 O 表示)和反应变量(reaction variable/response variable，常以 R 表示)。心理学的实验研究就是要探明这三类变量间的相互关系，特别是因果关系，因此现代心理学的实验方程式可以写成如下形式，它表示人的心理或行为改变是刺激变量与机体变量共同作用的结果：

$$R=\mathrm{f}(S,\ O)$$

在具体研究中，研究者就是要考虑对这三类变量进行控制、操纵或观测，以探究其中是否存在某些因果关系。

1. 刺激变量

刺激变量来自外部的刺激情境，是心理实验中对被试的心理或行为可能产生影响的各种外在条件或因素。在一项心理实验过程中，可能的刺激变量很多，

① 在国外有些同类教材中，或一些作者的研究报告中，对人类和非人类被试做了区分。人类被试叫做“participant”，非人类被试叫做“subject”。但在行文表述中又常常发生混淆，所以国内作者一般不对其做区分，不管是人类被试还是非人类被试，都叫做“被试”(subject)。

如环境光线、声响、人际交互作用等等。但是一项心理实验不可能对所有刺激变量的影响都进行研究,只能根据研究的目的有选择地考察一个或几个变量对被试反应的影响,为此又必须同时对其他的刺激变量进行控制,防止其成为实验中的混淆变量。所以,刺激变量又可被分为两类:一类是研究者拟研究并要借助实验过程以探明其对被试的心理或行为能否产生影响的变量,叫做自变量或独立变量(independent variable)。自变量是主试在实验中要进行操纵的、会按照主试意图有步骤变化的变量。除自变量以外,还有另外一类刺激变量不是主试拟研究但也可能对被试的心理或行为产生影响,这些变量叫额外变量(extraneous variable),它是主试要控制的,以防止其成为混淆变量,因此也叫做控制变量(controlled variable)。

所谓混淆变量(confounding variable),是指那些不拟研究但却与自变量一起发生系统性改变的变量,它使得被试心理或行为变化的原因变得无法确定,故称为混淆变量。

例如,在一项关于灯光刺激的颜色对简单反应时间影响的实验研究中,灯光刺激的颜色、强度、面积、持续时间以及环境噪声、主试特征等都是刺激变量,在这些刺激变量中只有灯光颜色是自变量,其他刺激变量均为额外变量。主试必须操纵自变量,让其按照研究者的意图有步骤地发生改变,比如对一组被试施加红光刺激并测定简单反应时间,对另一组被试施加绿光刺激并测定简单反应时间。但是,如果在自变量改变时,某一额外变量未被控制,也与自变量一起变化,它就会成为混淆变量。比如,实验中,主试 A 主持了红光刺激条件下简单反应时间的测定后因事离开了实验室,于是由主试 B 接替 A 的工作,继续主持绿光刺激条件下简单反应时间的测定。结果两种条件下反应时间出现了差异,那么这种差异是由灯光颜色改变引起的,还是由主试改变引起的呢? 由于主试特征这一额外变量与灯光刺激颜色这一自变量一同改变,就使得实验中观测到的被试的反应差异难以归因了,这就发生了变量混淆。为避免混淆,必须尽可能地控制自变量以外的其他影响变量,使其不与自变量一起发生系统性改变。

2. 机体变量

机体变量是指可能对被试的心理或行为发生影响的被试自身的特征或身心状态。如被试年龄、性别、身心健康水平、教育水平、特殊训练、动机、性格、内驱力强度等,都是常见的对被试的某种反应可能产生影响的变量。在机体变量中,也同样包括研究者拟研究的和不拟研究的,所以机体变量也可以分为两类:自变

量和额外变量。但是,当把机体变量作为自变量时,它并不是完全意义上的自变量,因为机体变量是被试固有的或相对稳定的特征,并非由研究者或主试随意操纵。不少教科书把这样的自变量叫做准自变量(quasi-independent variable)。但从另一方面来说,这类变量虽然是研究者不能随意操纵的,但研究者可以按照实验设计的要求主动选择机体变量的水平并将其作为分组变量。许多情况下,不对准自变量与自变量作严格区分,笼统地都叫做自变量,除非有特别的要求。

机体变量中,除拟被研究的准自变量或自变量外,也存在一些可能会成为混淆变量的机体变量,即额外变量,这些变量也是实验过程中需要加以控制的。如一项研究欲考察性别与形象思维能力、抽象思维能力发展水平的关系,性别就是准自变量,年龄、教育年限等就是需要控制的额外变量。

3. 反应变量

反应变量是指由于自变量的变化而引起的被试反应或内外变化,所以也叫因变量(dependent variable)。反应变量是在实验中需要观测和记录的变量,通常包括反应的速度、强度、难度、准确度和频数等,如不同光照条件下的反应时间,这是反应速度;不同刺激情境下,皮电测试仪指针偏转的读数,这是反应的强度;智力测验中,完成的作业等级,这是反应的难度;走迷宫实验中,完成一轮操作走入盲项的次数,这是反应的准确度;不同教育方式下,学生利他行为的次数,这是反应的频数。这些变量,都是易于观测和记录的变量,是实验中常用的因变量。

在上述变量中,我们注意到:自变量是研究者操纵或选择的,因变量是研究者要观测并记录的,这样做的目的就是要查明:因变量是否随着自变量的变化而变化。如果因变量随着自变量的变化而变化,则二者很可能存在因果关系,否则就不具有因果关系。不过,这样的推断需要一个条件,即对实验中的额外变量进行控制,使之不沦为混淆变量。

根据以上的变量分析,我们可以把各类变量的关系表示成图 1-1 的简单形式。

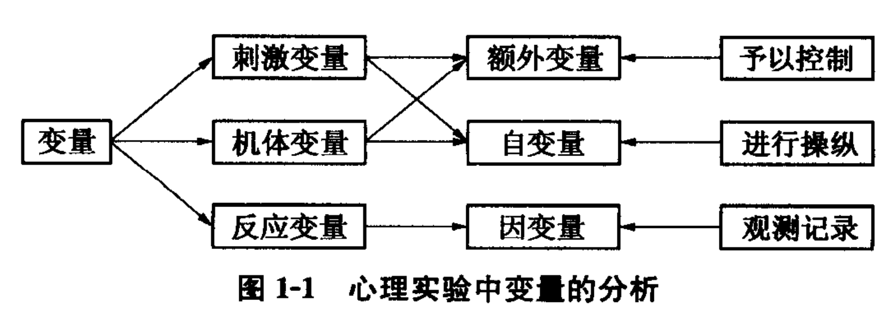


图 1-1 心理实验中变量的分析

(二) 因变量的基本条件

在心理实验中,因变量对于探测变量间的因果关系至关重要。为实现实验目的,实验中的因变量必须稳定、有效,具有较强的区分能力,即要具有信度、效度和区分度,这是因变量应具备的三个基本条件。

1. 信度

信度(reliability)也叫做测量的一致性,对于因变量来说,就是其在相同或相近条件下测量结果的前后一致性。比如,一个被试参加灯光刺激简单反应时间的测试,测试的条件保持不变,但是测试的结果前后相差很大,分布在 50 毫秒～520 毫秒之间,显然因变量测量的结果前后不具有一致性,这次测量得到的变量值就不能作为因变量,因为这样的测量是不可信的,也就是不具有信度。不具有信度的因变量无法用于自变量效应的检测,因此也不能用于对自变量与因变量间因果关系的推断。相反,如果测试的结果在某一个值附近有所变化,但基本一致,则测量是可信的,具有测量的信度。

那么如何判断一项因变量测量的信度高低呢? 简单地说,就是看测量条件不变时多次重复测量的结果的变异性大小。比如,多次参加智力测量时得到的智商分别为 120、85、105……信度就很低;如果得到的智商分别是 120、110、127……测量信度就较高。重复测量某一特定声音刺激条件下的简单反应时间,得到的结果分别为 150 毫秒、180 毫秒、89 毫秒、320 毫秒、35 毫秒……测量的信度也很低,数据不可用。像心理量表测量中的折半信度、重测信度、内部一致性等也都是在考察测量的变异性大小。变异性较小则信度就较高。但是,信度高不代表测量就一定可靠,它还必须具有较高的效度。

2. 效度

效度(validity),就是测量的有效性,即因变量能不能确实测量到它所要测量的被试心理或行为的变化。例如,随着作业难度的增加,测得被试在相同时间里识记的项目数减少了,这种结果能够有效地反映作业难度对识记的影响,就是有效的。如果在测量中,较难的识记材料识记的项目数更多,这测量就不能正确地反映作业难度对识记的影响,其中的原因可能是那些较难的实验材料是他曾经学过的,或他在学习这一材料中投入的心理资源更多。再如,有人用磅秤测量人的身高,虽然有信度,但却没有效度。

当测量信度较低时,自然就没有了高效度。但是当信度较高时,又如何判断这项因变量测量的效度高低呢? 有时,这种评估很容易,因为它是“显而易见”

的，就如上述用磅秤测量身高的例子，其效度显然是不高的。有时，这种评估就较为困难，因为它不能仅从数据变化来判断。除了可以使用同类测量进行佐证之外，还可以聘请专家来评估，即所谓“专家效度”(expert validity)。为什么专家有资格对测量效度作出评判呢？因为专家理解你所要测量的心理现象或心理品质的本质，以及这些心理现象和心理品质与外部反应之间的关系。比如，一项智力测量是不是真的能测量人的智力水平，这需要真正理解智力本质的专家才能评估。

需要特别注意的是，当测量的因变量不具有高效度时，研究可能得到错误的、甚至荒谬的结论。这种错误或荒谬即使没有被专家评审时剔除，也最终会在实践中暴露出来。

3. **区分度**

随着自变量的变化，因变量也随着变化，即因变量对自变量的变化比较敏感，我们说这种因变量具有区分度(discrimination)。比如，在记忆能力的测试中，如果测验过于容易或给予被试的学习时间太多，所有被试都取得很高分数，显示不出被试间本来有的记忆能力差异，这些分数就不具有区分度；或者，在操纵实验中，记忆任务较小或给予被试学习的时间太多，不同处理条件下被试的学习成绩都很高，也不能区分出不同处理条件对记忆和学习的影响，这种现象叫做“天花板效应”(ceiling effect)。相反，如果学习任务太难或给予被试的学习时间太短，被试的分数都很低，这些分数也不能反映被试间记忆能力的差异，或不能显示各种处理条件对记忆和学习的影响，这样的因变量也不具有区分度，这种现象叫做“地板效应”(floor effect)。

因变量必须同时具备以上三个基本条件，才是较好的心理实验中被试心理或行为的有效测量指标。一般来说，信度和区分度是效度的必要条件，但不是充分条件。当信度和区分度较高时，因变量才可能具有高效度；具有较高效度的因变量则肯定具有较高的信度和一定的区分度。就严格的心理实验来讲，因变量的信度和效度分析是必要的，除非它是“显而易见”的。

(三) 额外变量的控制

如前所述，心理实验的主要目标是探明自变量与因变量之间的因果关系，但如果额外变量与自变量一起系统变化就会成为混淆变量，造成因变量的变化出现归因困难，因此心理实验成功的关键在于能否尽可能地消除额外变量的影响，即对额外变量进行控制。进行有效控制，首先要分析一项实验中可能存在的额

外变量,这往往需要研究有关的文献资料,进行一些预备实验、掌握一些理论基础知识、根据经验进行估计,并在预备或正规实验中处处留心,一旦发现额外变量,及时考虑进行控制。常用的控制方法有以下几种。

1. **消除法**

对于有些可能影响被试心理或行为的因素,可采取消除法,即避免该因素在实验过程中出现。最常见的是对环境光线或环境噪声影响的处理,方法是将实验安排在暗室或隔音室中进行。消除法的优点在于将某些额外变量的影响完全隔离,缺点是会造成研究情境与现实情境的较大差异,引起研究外部效度的下降。所谓外部效度(external validity),就是研究结果或结论的概括性或可推广性。研究情境越具有一般性、越接近于现实情境,得到的结果或结论就越能说明心理生活的真实性,因此也越具有可推广性,其外部效度就较高。否则,外部效度就低。

2. **恒定法**

有些额外变量无法消除或没有消除它的条件,则可以使其保持恒定,即在整个实验进程中使其保持在某一恒定水平,以保证其对所有被试的影响基本一致。比如实验中如果没有暗室可用,环境光线对实验可能产生影响,就可以采用恒定法,安排所有被试在同一个实验室中完成实验任务,而且使实验室中的照明条件保持恒定;还比如,如果一个实验无法在一次完成,那么在安排时间上也要尽量保持一致,如所有被试都安排在每天下午的 3 点到 5 点参加实验。

3. **匹配法**

在分组实验如实验组-控制组实验中,为消除被试组差异对实验结果的影响,要尽可能做到等组实验,即两组或多组被试在有关因素上的平均水平保持一致,这时可采用匹配法。这种方法首先就是要对所有被试的有关因素进行初测,根据初测的结果来对被试进行匹配性分组。匹配中不仅要使各组被试在该因素得分的平均数几乎相等,而且要保证其标准差也几乎相等。如为了研究问题的不同呈现方式是否会引起问题表征难度的变化,要安排两组被试分别在两种呈现方式下解决结构相同的问题。考虑到被试的智力差异可能会造成变量混淆,需要对两组被试进行智力匹配以保证两组被试在智力水平上相当。于是分组前先对抽取来的 20 名被试进行智商测试。假设的测试结果为:

108　127　98　85　120　115　105　96　100　90

95　115　114　120　109　95　125　123　108　119

如何进行匹配呢？首先将这20人的智商分数按大小排序，然后两两配对，再将每一个配对的两人分别分到两个组：85、***90***、***95***、95、96、***98***、***100***、105、108、***108***、***109***、114、115、***115***、***119***、120、120、***123***、***125***、127。分组时要注意不能将每一对中分数高的或分数低的都分到一个组。具体的分组结果是：

A组：85、95、96、105、108、114、115、120、120、127

平均数和标准差：108.50±13.26

B组：90、95、98、100、108、109、115、119、123、125

平均数和标准差：108.20±12.21

由于两个被试组在智力水平上取得了较好匹配（平均值与标准差都很接近），智力因素就不会成为实验中的混淆因子，有利于提高研究的内部效度。所谓研究的内部效度（internal validity），就是因变量变化原因的确定性。当自变量以外的其他变量得到比较好的控制或平衡后，实验中较好地消除了混淆变量，因变量的变化基本上都可归因于自变量的变化，实验就具有较高的内部效度；如果因变量的变化除可能是自变量的变化引起之外，也有可能是其他未被很好控制的额外变量引起的，研究的内部效度就降低了。上述例子中，如果两组被试智力水平有较大差异，他们在实验中作业成绩的差异既有可能是问题呈现方式不同引起的，也有可能是智力水平不同引起的，归因就出现了困难，内部效度就下降了。

4. 抵消平衡法

这种方法常用于克服实验中的空间位置效应和时间效应。在心理学实验中，空间位置效应往往是指刺激呈现的位置或被试操作键的位置引起的因变量的变化；时间效应也叫做顺序效应（order effect），一般包括练习效应（practice effect）和疲劳效应（fatigue effect）两种。比如一个被试在一项研究中要做两种实验A和B，最后要对其两种实验的结果进行比较，于是其实验的顺序可以为：ABBAABBA……或BAABBAAB……或ABBABAAB……当然，这种安排有一个前提，即顺序效应呈线性变化。如果是非线性的，则要采用更复杂的抵消方法，此处不作介绍。这种方法既可用于同一被试参加不同实验的实验顺序安排，也可以用于不同被试（A组、B组）参加同一实验时实验顺序的安排。

5. 随机化法

就是按照随机化的方法抽取被试、分配被试或编排实验，不附加任何别的条件。如在50人中抽取10人参加实验，这50人中的每一个人被抽取的概率都是

1/5。具体方法可以是抽签法、随机数表法、随机排队法等。除被试的选取和分组可用随机化法以外，实验顺序的编排也可用随机化法。采用随机化方法可以对许多偶然因素对实验结果的影响进行控制和平衡，特别是能有效地平衡各被试组间机体变量的差异性。

概括上述关于变量的分析，可以认为心理实验的核心本质在于：主试操纵自变量、控制额外变量、观察记录因变量，以揭示自变量与因变量间的因果关系。不过，这里需要说明的是，我们通常所称的自变量包括三种情况：第一种情况是，自变量属于刺激变量，如灯光刺激的强度、声音刺激的频率、实验环境中的气温等；第二种情况是，自变量属于由刺激变量引起的机体变量，如由不同的奖励方式引起被试不同的动机强度、不同的环境刺激引起被试不同的情绪状态等；第三种情况是，自变量是被试固有的机体特征，如被试的性别、年龄、身高等。前两类自变量是可以操纵的，研究者可以根据研究需要有步骤地改变它，以探测其与被试心理和行为改变的因果关系。第三类自变量不是研究者可以操纵的，所以它不是真正意义上的自变量，有研究者将其称为准自变量(quasi-independent variable)。① 研究者可以利用准自变量对被试进行分组，以探测其与因变量的共变关系，如果二者真的存在共变关系，这种关系也未必就是因果关系，可能只是一种相关关系，除非研究者还有其他方面的论据。比如，测量发现，身高在 1.2 米左右的被试组与身高在 1.5 米左右的被试组存在显著的智力差异，这里存在的就只是一种相关关系，肯定不是因果关系。本书的表述中不再区分自变量和准自变量，因为国内外许多刊物上发表的研究报告也都未对此加以区分。在此给予特别的说明，是为了提醒读者在阅读研究报告时需注意这一点，在研究中要作出因果关系结论时更要注意这一点，尽量避免错误地将相关关系看作因果关系。

二、被试的选取

每一项实验研究都是一个独立事件，它需要一个或多个具体的被试组。但大部分研究试图要解答的都是关于较大群体的一般问题，而不是关于较小群体的、少数特定人的问题。因此，研究者一般都期望将他们的研究结论推广到研究

① [美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著，邓铸、姜子云、蒋小慧等译：《行为科学研究方法》，陕西师范大学出版社 2005 年版，第 165 页。

被试之外的范围。那么,这其中就存在一对矛盾:一方面要选取较少被试参加实验,另一方面又期望将从这少量被试得到的研究结论推广到一个大的群体。这一矛盾如何解决呢?关键是要使选取来的少量被试能够具有代表性。

在介绍被试选取方法之前,需先澄清几个概念:目标总体(target population)、可得总体(accessible population)、代表性样本(representative sample)、有偏样本(biased sample)等。任何确定了的课题,在进入设计阶段时,都要首先根据研究目的确定研究对象,研究对象的全体构成了研究的目标总体,这目标总体的所有个体应该具有某种共同特征。比如,研究某种药物或治疗程序对抑郁症患者是否有效,那么目标总体就是全体抑郁症患者;研究贫困地区学龄儿童的自我意识发展,则其目标总体就是所有贫困地区的学龄儿童。这里,研究者要对"贫困地区"进行界定。不过,实际研究中很少有研究者真正能从目标总体中来选取被试,因为目标总体常常不容易得到。要研究药物对抑郁症患者的疗效,其目标总体应该是世界上所有抑郁症患者,可是你有可能到世界各地去选择被试吗?或者范围再缩小些,把全国的抑郁症患者作为目标总体,那么你可能到全国各地去选择被试吗?几乎肯定是不可能的。最有可能的是在你所在的城市及其附近,去访问一些心理咨询与治疗机构,也包括一些精神治疗机构、综合医院精神科门诊,从这里所能找到的抑郁症患者就是你可能用为被试的研究对象,叫做可得总体。大部分研究者都是从可得总体中选取样本的,因此不仅需要在将研究结果推广到可得总体时谨慎从事,而且在将结果推广到研究的目标总体时更要格外小心。我们在此后讨论中,所谈的"总体"主要是指"可得总体"。

为使研究结果能被推广到总体,选取的样本就要具有代表性,即形成代表性样本。所谓代表性样本,就是在与研究有关的特征方面,样本与总体一致(误差在允许范围内)。相反,如果样本特征与总体特征相差甚远,超出了误差许可的范围,这样的样本就叫做有偏样本。在被试选择中,尽量得到代表性样本,避免有偏样本的出现。

需要指出的是,不管采取何种方法,从一个总体中抽取样本,误差总是存在的,所以样本特征与总体特征必然存在差异。而且这种差异符合统计学规律,即如果进行许多次抽样,抽样的误差分布往往符合某种统计学分布规律。因此,所谓代表性样本是指在统计学意义上该样本能代表总体。那么如何进行被试选取才能保证得到代表性样本呢?行为科学研究中,样本选取的方法包括概率抽样

(probability sampling)和非概率抽样(nonprobability sampling)两大类。其中概率抽样主要包括简单随机抽样、分层随机抽样、按比例分层随机抽样、整群抽样；非概率抽样主要是便利抽样。①

(一) 简单随机抽样

简单随机抽样(simple random sampling)的基本要求就是总体中的每一个体具有相等且独立的被选中概率，概率相等意味着任何个体都不比其他个体更有可能被选中，相互独立则意味着某一个体的被选择不会影响研究者对另一个体的选择。

简单随机抽样的过程一般包括三步：

第一步：确定一个总体，即你预备要从中选取样本的总体。该总体应该是一个可得总体。

第二步：列出总体中的所有成员，形成一个个体表列。通常，对表列中的所有个体编号。

第三步：根据研究需要，使用随机过程从表列中选择出一定数量的个体。这里所讲的随机过程可以是“抽签法”、“随机数字表法”等。

抽签法，是先将总体中的每一个体编号，并把每个个体的号码写在一张纸条上，再将纸条搓成团，混在一起并摇匀，最后随机捡出若干纸团，这些纸团上的编号就是被选取的被试的编号。

“随机数字表”就是由 0～9 的数字随机排列构成的数码表，它以 5 个数字为一组，如图 1-2 所示就是随机数表的一个小片段。一般的统计学教材或实验心理学教材都会将“随机数字表”作为附录。随机数字表法的操作程序是：先将被试编号，然后随机地从“随机数字表”中划出一个数表片段，从该片段的开始部分依次向后或向下搜索，当遇到一组数字的后边几位正好与某一个体的编号相同，就将该个体作为被试选出，依此方法继续进行，直到选够所需要的被试数为止。比如，要想从 100 人的总体中抽取一 20 人的样本，可以先将这 100 名个体编成 00～99 号，然后从数表中随机选择一个片段，如图 1-2 中第 5 到第 9 行、第 1 到第 7 栏，接着按顺序选号，这里可以选到的编号是：12、18、55、70、51、41、82、42、81、39、72、97、47、61、59、16、23、09、99、40，于是就构成了一个 20 人的样本。

① [美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著，邓铸、姜子云、蒋小慧等译：《行为科学研究方法》，陕西师范大学出版社 2005 年版，第 85—92 页。

23157	54859	01837	25993	76249	70886	95230	36744
05545	55043	10537	43508	90611	83744	10962	21343
14871	60350	32404	36223	50051	00322	11543	80834
38976	74951	94051	75853	78805	90194	32428	71695
97312	61718	99755	30870	94251	25841	54882	10513
11742	69381	44339	30872	32797	33118	22647	06850
43361	28859	11016	45623	93009	00499	43640	74036
93806	20478	38268	04491	55751	18932	58475	52571
49540	13181	08429	84187	69538	29661	77738	09527
36768	72638	37948	21569	41959	68670	45274	83880

图 1-2　随机数字表中的部分随机数

简单随机抽样从理论或逻辑上排除了选择偏好，一般可以得到代表性样本。但是，需要注意的是，简单随机抽样是通过把每一次的选择都置于随机性的规则之下来消除偏好的，它可以在较长的抽样过程中得到很好的代表性样本，就如投掷几千次硬币，最后的结果会是正面朝上和反面朝上各占 50%。但如果抽样过程较短，就有可能得到有严重偏向的样本，就像投掷一种质地均匀的硬币 10 次，甚至会出现 10 次都是正面朝上的结果；从 100 名女生和 100 名男生组成的总体中随机抽选 10 人，甚至会出现抽取的 10 人全为男生或全为女生的情况。要避免这种非代表性样本，研究者常常采用分层随机抽样和按比例分层随机抽样方法。

（二）分层随机抽样和按比例分层随机抽样

一个总体通常可以区分出各种不同的子群(subgroup)。如一所大学里的学生可以分为不同年级的子群、不同专业的子群、不同性别的子群等等。要保证在一个样本中，各子群都能得到代表，可以使用分层随机抽样(stratified random sampling)。为得到代表性样本，首先要确认样本中应包括哪些具体的子群或层，然后使用与简单随机抽样完全一样的步骤从每个预先确认的子群中选择相等的子群随机样本，最后把这些子群样本合并成一个总的样本。比如，我们计划从某学院的研究生中抽取 50 人的样本，要求其对男生和女生都具有代表性，那么可以首先从男生中随机抽取一个 25 人的样本，再从女生中随机抽取一个 25 人的样本，最后将这两个子群样本合并起来就构成了想要的分层随机样本。

当研究者想对总体中的各个部分进行描述，或对各个部分进行比较时，分层随机抽样方法就显得特别有用。采用这种方法，样本中的每一子群必须包含足

够的个体以便它能代表总体中与其对应的部分。

当研究焦点集中到总体中某一特定子群时，最好采用分层随机抽样方法来选择被试。也就是说，当研究要考察各具体子群并对他们进行比较时，这种方法是比较适当的。但如果研究的目标是考察整个总体，这种抽样技术可能会带来问题。最典型的情形是，总体中每个子群的实际人数不相等，但样本中各子群的代表人数都相等。比方说，在一个总体中某一子群的人数只占总体的3%，但它在样本中却占到25%的分量。克服这一问题的方法是采用按比例分层随机抽样，具体做法是：首先区分出总体中的各个子群或成分，并确定总体中相应子群所占的比例，然后根据计划的样本容量和各子群在总体中的比例数确定每一子群应抽取的被试数，最后从每一子群中抽取相应的被试数，合并在一起就可以得到一个其比例关系与总体中的比例关系完全匹配的样本。这种抽样就叫做按比例分层随机抽样(proportionate stratified random sampling)，或简称为按比例随机抽样(proportionate random sampling)。

（三）整群抽样

一般情况下，研究者都是通过从总体中选择个体而得到样本，但有时个体是以现成整群形式存在，研究者就可以随机地选择整组。比如研究者想从某个城市的学校系统中抽取一个由300名三年级学生组成的大样本，他不是一次选择一个学生，而是随机地选择了10个班(每个班大约有30名学生)，最后也可得到300人的样本，这一程序就叫整群抽样(cluster sampling)。只要在感兴趣的总体中存在有界定清楚的整群，就可以使用这一程序。这种技术有两个明显优点：第一，它相对快捷，容易得到大样本；第二，对被试的处理和测量常以整群方式进行，可以大大加快研究进程。在整群抽样中，研究者不是选择单个被试、不是对单个被试施加处理、不是每次只测量到一个分数，而常常是针对整群施加处理，每次可检测一群人，从一次实验中就能很便利地取得30个被试的数据。

（四）便利抽样

便利抽样(convenience sampling)是一种非概率性抽样方法，这是心理学研究中实际上最有用的抽样方法。在便利抽样中，研究者只使用那些容易得到的个体作被试，被选的人必须是那些找得到的、乐于参加研究的。所以，我们看到，在心理学研究中，使用大学生作被试是最常见的，因为这些学生通常就是研究者的学生。

便利抽样被看作是一种比较弱的抽样方法，因为研究者不试图去了解总体，

在选择被试时也不使用随机过程，对样本的代表性很少控制，所以得到有偏样本的可能性很大。像人们积极参加的广播电台听众热线电话调查或杂志社使用通信方式进行的调查都是特别值得怀疑的，这些情况下的调查样本应该是存在偏差的，因为只有那些倾向于收听这个电台节目或倾向于阅读这个杂志又对调查的主题感兴趣的人才愿意去花费这些时间，这些人不可能是一般人群的代表。

尽管存在这样大的缺点，但是便利抽样可能还是被使用最多的方法。与那些需要详细了解总体中所有成员情况、又需要采用费时费力的随机过程来选择被试的方法相比，便利抽样更容易、更廉价、更快捷。便利抽样虽然不能保证总能得到有代表性的无偏样本，但也不能就草率地将其看作是一种毫无补救希望的抽样方法，大部分研究者都知道，可以使用两种策略来纠正便利抽样中的主要问题。首先，研究者尽可能地确保他们的样本具有相当的代表性而无大的偏差；其次，详细地说明样本是如何得到的、参加研究的被试是哪些人。

在讨论被试选择时，我们还不得不涉及另外一个问题：心理学研究中的伦理学问题。为了遵循研究者的伦理规则，在选取被试时必须注意以下几点：(1)研究中必须保证被试的安全，尽可能使之免遭心理或生理的伤害。如果研究无法避免这样的伤害，那也必须是轻微的、事后可补救的，而且在使用这些导致轻微伤害之前，需要向拟作为被试的那些人作出说明，征得同意后才可实施。实验结束后，要向被试作出说明和进行必要的安抚，以消除实验的消极影响。(2)研究中尽可能地避免依靠欺骗手段或强制手段，让一些人参加实验，尤其是那些对被试有潜在伤害性的实验。前些时候媒体报道，某国外医疗机构打着援助我国某一经济落后农村的旗号，在我国偏远农村进行某项药物实验，实验中隐瞒了该药物的潜在危害。这种做法是非常不人道的。在心理学实验中要避免类似的情况发生。(3)由于实验中，选取被试存在实际上的困难，不能保证得到无偏样本，但在研究报告的撰写中，对此加以“掩饰”，有意错误地表述样本的抽取过程。这一做法会干扰读者对研究结果优劣的评判，也是不允许的。

最后，还需要说明的是：如何确定样本容量的大小。样本容量大小没有绝对的标准，也不存在严格的计算方法，但依据研究本身的特点和目的而言，确定样本容量实际上要在可行性与准确性之间进行平衡。一般来说，样本容量越大，结果准确性越好，但研究实施的难度越大；样本容量越小，结果准确性越差，但研究实施的难度越小。如何取舍，除考虑准确性外，还要看研究的内容与研究的类型。以下三个方面的考虑对于确定样本容量是有帮助的。

第一,研究的内容:研究中所要测量的心理现象或心理品质越受到生物性的制约,个体间的差异就越小,需要的研究样本就可以较小,如关于感知机制的研究、事件相关电位(ERP)变化模式的研究等;研究中所要测量的心理现象或心理品质越受到社会文化的制约,个体间的差异就越大,需要的研究样本就越大。

第二,研究对象总体中个体间的同质性:总体中个体间的同质性越高,个体差异越小,根据抽样规律,抽样误差也越小,需要的样本容量就可以较小;反之,需要的样本容量就较大。

第三,研究的类型:利用心理实验室严格控制实验条件,对被试的心理活动或心理特征进行观测,测量过程中产生的误差较小,研究样本可以较小;利用自陈量表对被试的心理特征进行测量,被试反应容易受到多种因素的影响,产生的测量误差会比较大,研究样本就需要较大。

三、实验设计模式的选择

课题选定后,也清楚了实验中要控制、操纵和测量的变量,研究被试如何选择也非常清楚了,那么实验如何进行才能保证研究目的的实现呢?在心理学长期实验研究中,逐渐形成了一些稳定的实验编排程序和策略,我们将其叫做实验设计模式(pattern of experiment design)。分析这些设计模式,可以发现其中最具一般性的基本实验设计类型是组间设计、组内设计和混合设计,其他更复杂的实验设计模式都是在这些设计类型基础上演变出来的。为了更清楚地表述心理学实验研究的程序,我们在此先介绍三种基本的实验设计模式,其他实验设计模式留待下一章专门介绍。

(一) 组间设计

组间设计(between-groups design),也叫做被试间实验设计(between-subjects design),其基本原理是:将被试随机分配到不同的自变量或自变量的不同水平上,各自独立地在不同的处理条件下接受因变量的测量。就是说,要把被试随机分成几组,每一组只参加一种条件下的实验,组与组之间的反应结果不产生相互影响,因此是独立的,所以这种实验设计也被称为独立组设计(independent groups design)。在这样的实验设计中,其难度或关键是被试组之间差异的克服,也就是要尽量做到参加实验的各组被试间基本上是相等的。这样,就可以观察接受不同实验处理的各组被试在有关因变量上是否有差异,如果存在差异,这

差异应该就是由实验处理的不同造成的。以声光刺激简单反应时间的比较为例来说明之。

研究者为了研究声光刺激条件下，人的简单反应时间是否存在差异，就从一所大学的某一班级随机抽取了 20 名男大学生，接着又随机地将他们分成了相等的 A、B 两组，每组 10 人。A 组被试参加声音刺激条件下的简单反应时间的测定、B 组被试参加光刺激条件下的简单反应时间的测定，两组被试实验的其他条件完全一样。每个被试测试 20 次，并将 20 次平均，得到每个人的简单反应时间。假设得到如表 1-1 所示的测量数据。

表 1-1　声光刺激简单反应时间的比较

被　试	1	2	3	4	5	6	7	8	9	10
声反应时间(ms)	150	163	157	171	170	165	160	185	156	160
光反应时间(ms)	189	202	210	220	193	195	232	196	185	200

得到这样的实验数据后，需要采用统计学方法分别计算两组被试反应时间的平均数和标准差，并通过统计检验(t 检验)确定两组被试反应时间是否存在显著性差异(这些数据分析方法详见附录)。如果两组被试反应时间存在显著性差异，则说明不同感觉通道的刺激引起反应的速度是不同的；如果两组被试的反应时间没有显著性差异，则说明声、光刺激的不同条件下，被试的反应速度没有明显差异。根据这一研究，可以验证刺激通道与反应速度之间是否存在因果关系。

除考察两种实验条件下被试反应是否存在差异外，组间设计也包括超过两种实验条件的情况。例如有研究者想考察不同奖励方式与员工工作绩效的关系。他设计三种奖励方式：将奖金加到薪水里到月底评估和发放一次、每周评估和发放一次、每 10 天评估和发放一次。以半年内平均每周完成的产品件数作为工作绩效测量。将 300 名员工随机分成三组，分别到三个车间从事同样的工作，工作环境和其他环境因素、接受的管理等都一样，只是奖金发放方式不同。这一设计也是组间设计，其数据需要通过方差分析进行处理(具体数据分析过程见第二章或附录 1)。

组间设计的优点在于，不同处理条件下的实验由不同被试组各自独立完成，得到的各组数据互不关联，相互独立，实验中也不存在各种处理条件间的相互干扰，同时每一被试的实验任务相对较少，易于完成。缺点在于，需要的被试数较多，而且如果被试分组出现偏差，则会出现各组间被试的差异，降低研究的内部

效度。因此,如果能抽取来的被试数较少,或者被试间的差异性较大,而且实验中顺序效应不太明显时,可以考虑采用组内设计。

(二) 组内设计

组内设计(within-group design),也叫做被试内设计(within-subjects design),是把抽取来的所有被试作为一组,接受自变量的多种处理,也就是说所有被试要在实验中自变量的所有处理水平上重复进行实验。比如,前述的声光反应时间比较的实验也可以采取组内设计,就是让一组被试既在光刺激下测定简单反应时间,还要在声刺激下再测简单反应时间,所以这种实验设计也被称为重复测量实验设计(repeated measures design)。这种实验设计的主要问题是一种实验条件下的操作会影响被试在另一种实验条件下的操作,即所谓的顺序效应(order effect)。但这种设计节省被试,而且被试组的差异对各种实验条件来说是不存在的。为了解决顺序效应的问题,就要在实验顺序的安排上采用抵消平衡方法。我们以下列实例来说明。

某研究者想通过实验证实缪勒错觉,于是抽取了 10 名被试,每个被试都先后用长度估计测量器测量长度估计误、用缪勒错觉仪测量缪勒错觉量。两项实验都重复若干次,即可得到 10 名被试长度估计误和缪勒错觉量的平均值,如表 1-2 所示。在实验的操作过程上,要特别注意顺序效应的影响,需要采用平衡法避免顺序效应成为混淆因子。具体做法是:每个被试在进行“长度估计误”(A)测量和“缪勒错觉量”(B)测量的顺序是:A-B-B-A-A-B-B-A……或者是:B-A-A-B-B-A-A-B……最后得到每个被试两种测量的结果。

表 1-2　长度估计误与缪勒错觉量的比较(mm)

被试序号	长度估计误	缪勒错觉量
1	2	6
2	1	5
3	2	7
4	3	9
5	4	10
6	2	11
7	3	9
8	2	7
9	1	5
10	2	8

这一实验设计得到的数据也需要借助统计学方法进行分析。首先得到每组数据的平均值和标准差,然后使用相关数据的差异 t 检验方法来确定其差异是否显著(详见附录1)。如果差异显著,则说明同一组被试在两种测量条件下的长度估计误差不同,证明缪勒错觉已经不同于一般的长度估计误,所以缪勒错觉现象存在;如果差异不显著,则说明所谓的缪勒错觉实际上与一般的长度估计误没有什么不同,缪勒错觉是不存在的。

同样,组内设计也可以考察两种以上处理条件下反应的差异性,比如在选择反应时间测量中,测量条件包括"红、绿、蓝、黄"四种颜色的灯光刺激。可以选择一组被试参加选择反应时间的测试,测试中不同颜色的灯光刺激随机呈现,最后比较各种颜色刺激下选择反应时间有无差异。当数据超过三组时,数据分析也要采用方差分析,但这种设计的方差分析与组间设计的方差分析有所不同(详见附录1)。

组内设计的优点在于,不同处理条件下的实验均由同一组被试完成,被试差异不会成为混淆因子,研究的内部效度较高,而且这种设计较为节省被试。缺点在于,组内设计中不同处理条件下的实验并不独立,容易出现相互干扰和顺序效应,被试的实验任务往往较大。一般,在被试相对较为缺乏,不同处理间的干扰较小、顺序效应较小时采用组内设计,但要注意顺序效应的抵消和平衡。

(三) 混合实验设计

以上的组间设计和组内设计各有优缺点,如果把两种设计结合起来,则可达到相互弥补的效果。以实例说明之:现有20名被试参加大小常性系数测试,要求在实验中考察距离不同、单眼和双眼观察时大小常性系数是否不同。实验仪器是大小常性系数测量器。

这一实验有四种不同的实验条件,如果采用组间设计,则需要将20名被试分成4组,每组只有5人,样本容量太小,实验误差就会较大;如果采用组内设计,每个被试要在四种条件下接受测试,实验任务太大、前后干扰也较大。所以,本研究宜采用混合设计:把20名被试随机分成两组,每组被试10人。一组被试采用单眼观察,在4 m、8 m两种距离上测试大小知觉常性系数;一组被试采用双眼观察,在4 m、8 m两种距离上测试大小知觉常性系数,这就是混合设计。对于自变量"观察方式"(A)来说,两组被试是独立的,属于组间设计;对于自变量"观察距离"(B)来说,实验设计是组内设计。混合设计能够取得较为丰富的实验资料,也是最常用的设计方法之一。其数据可以采用多因素方差分析来进

行统计检验(详见附录1)。这一实验设计的模式和假想的数据如表1-3所示。

表1-3 混合实验设计模型

被试编号	单眼观察(A_1)		双眼观察(A_2)	
	4 m(B_1)	8 m(B_2)	4 m(B_1)	8 m(B_2)
1	0.52	0.32	0.92	0.62
2	0.35	0.35	0.75	0.55
3	0.45	0.25	0.85	0.75
4	0.52	0.42	0.82	0.72
5	0.35	0.30	0.75	0.75
6	0.55	0.45	0.85	0.85
7	0.65	0.45	0.75	0.65
8	0.55	0.33	0.85	0.70
9	0.55	0.25	0.95	0.75
10	0.54	0.44	0.74	0.78

一旦选定了实验设计的模式,一项实验设计的基本程序就比较清晰了,而且实验中取得的数据的结构和性质也都比较清楚了。接下来,可以依照实验设计所规定的实验任务准备相关的仪器与材料,做好实验前的一切准备工作。至此,实验就可以进入实施阶段。

第四节 心理实验的实施

实验设计只是完成了实验研究方案的制订,能否达到研究目标,还要看实验的实施。为了保证实验过程的顺利进行和较少出现各种不利的因素,研究者要在实验实施前做充分准备,规定实验时间、地点、仪器设备与材料(也包括实验用的纸笔等)、被试编排方式与实验顺序、数据记录表格等等。总而言之,实验的准备要"面面俱到"。

由于心理学的实验过程较为特殊,实验者和实验对象都是具有心理活动的人或动物,特别是人类被试具有的主观能动性往往会在实验中带来预想不到的混淆变量。如果不能有效控制实验进程,很容易导致实验失败或得出错误结论。

心理实验就是一个主试和被试相互作用的过程,①这种相互作用主要表现

① 黄希庭:《心理学实验指导》,人民教育出版社1988年版,第5—8页。

为三个方面:指示语效应、实验者效应和被试的能动效应。

一、指示语效应

这是主试与被试之间根据实验目的所发生的相互作用,是通过向被试交代任务而实现的主试对被试的直接干预。我们把主试向被试交代任务时所讲的话叫指示语(instruction),并把指示语对被试心理或行为发生的影响叫做指示语效应(instruction effect)。指示语在告知被试如何参加实验和完成实验操作的同时,也在将被试的心理活动引向有利于实验目的实现的方向,因此,它成为实验研究能否成功的关键因素之一。那么,如何制订指示语才能使实验紧紧围绕实验目的进行呢? 一般要注意以下几点:

第一,指示语要清楚、全面,避免使用专业术语。指示语要完备地、清晰地告知被试参加实验的任务、如何操作,要对被试行为作出明确规定。用语要避免使用专业术语,研究者在制订实验方案时往往会在这一点上出问题。有些词汇在主试看来非常简明、浅显,但对于未学习过心理学的人来说却是深奥难懂的。为了避免出现被试不能理解指示语的情况,除研究者要充分了解被试的知识水平外,还要在正式实验前,进行预备实验以检验实验程序及指示语的设计是否合适。

第二,指示语要简明扼要,不能使用模棱两可、一语双关的词汇,以避免难记和不同被试的不同理解。指示语不必对实验理由、目的、原理作过多解释,因为这种解释会加长指示语,造成难记,同时也增加了被试理解的负担,反倒使其不知道在实验中究竟要做些什么。也不能使用意义含糊、容易发生歧义的词汇,否则会造成不同被试的不同理解,因此导致不同被试以不同方式或不同的心理定向参加实验。

第三,指示语必须标准化,即主试或研究者在接触被试前要研究、讨论、确定指示语,然后将其写成书面材料或进行录音,以便在向被试呈现指示语时不出现差异性。这里的标准化包括指示语呈现的语气、声音高低,甚至发音者也应是同一人。当然,也不是所有的实验对指示语的标准化都有很高的要求,这要由具体实验的任务、要求和性质而定。

二、实验者效应

主试在实验中,不知不觉的期望、动机、疲劳、厌倦等心理活动,对被试会起

着一种颇为微妙的作用，如罗森塔尔效应(Rosenthal effect)①，这类效应叫做实验者效应(experimenter effect)。许多实验过程是主试和被试共同操作完成的，如果主试操作不熟练，也会对被试的心理活动或操作产生不利影响。心理实验中的主试要训练有素，持中立立场，善于把握实验进程，才能克服实验中的不良效应。

罗森塔尔效应：在罗森塔尔的一个研究中，他让选修心理实验课的学生用老鼠做实验，学生(实验者)得到用来做实验练习的老鼠，有些笼子上贴有“跑迷津伶俐”的标签，有些笼子贴有“跑迷津呆笨”的标签(其实，这些老鼠是随机分配到笼子里的)。于是，学生(实验者)根据笼子上的标签就误认为他们得到的老鼠是伶俐的或呆笨的。他们得出的实验结果是：标记为伶俐的老鼠比那些标记为呆笨的老鼠学习得快。实验者的期望莫名其妙地影响了老鼠的行为，动物的行为居然符合错误的标签。

——资料来源：《心理学实验指导》(黄希庭，1987)

三、被试的能动效应

人类被试具有主观能动性，并以此对主试及实验结果产生影响，它主要表现在三方面：

第一，人类被试往往不是被动地参加实验，而是和主试一样，企图在实验中满足自己的目的。有的被试会自己假定一个实验目的，然后在实验中进行验证；有的被试专门采取与主试对抗的态度参加实验；有的被试则故意迎合主试。实际上，这些态度都是与主试采取了不合作的态度，都不利于得到可靠的实验结果。

第二，实验中的霍桑效应(Hawthorne effect)和安慰剂效应(placebo effect)。所谓霍桑效应是指美国心理学家梅奥从1924～1932年间在霍桑公司的电力工厂开展的一系列霍桑实验中发现的一种现象：“即使在照明条件变坏和福利措施

---

① 罗森塔尔(Robert Rosenthal，1933—)德裔美籍教育心理学家，以研究教师期望效应(即“皮格马利翁效应”)而闻名。通过实验研究，证明教师对学生的不同期望能轻易地引起“自我实现预言效应”。1979年，获美国心理学会杰出专业贡献奖。

取消的情况下，实验组的生产量仍然上升，其原因就在于这些工人由于受到研究组和厂方的重视，使他们感到自己受到公司的重视，参与了公司重大问题的解决，从而产生了胜任感、成就感、自豪感。这些都激发了工人内在的积极性，从而提高了工作效率。"①在心理实验中，被试由于参加了实验，感到新奇、受重视，也会产生类似于霍桑效应的心理活动，进而影响实验目的的实现。

安慰剂效应是一种因为误认为服用了有效药物而产生的心理疗效。在心理实验中，被试也会对实验中可能的影响进行猜测，由此产生某些积极的心理效应。但这些效应本身不是实验处理引起的，所以反而会造成对处理效应的掩盖或夸大。总之，霍桑效应和安慰剂效应都不利于实验目的的实现，是研究者应注意避免的。

第三，被试影响主试，再转而影响被试自身。在实验中，被试的良好表现或较差表现、被试的迎合或逆反等会对主试产生影响，引起主试的满意、欣慰、赞赏、不满、焦躁甚至厌恶等心理或生理反应，这些反应又会反过来对被试产生影响，最后给实验结果带来消极影响。

心理实验中的关键因素是人的因素，其对实验的影响，有的是可以预见的，有的则无法预见；有的可以在实验中观察得到，有的则需要等实验结束后向被试询问才能知道。总之，心理实验中要时时想到被试不是被动的机器，要从多方面考虑控制条件，以及如何取得和分析与实验结果有关的材料。所以，一般在正式实验结束后，要把实验的原始结果收集、整理，保持完整，还要对被试做必要的访谈，以了解在实验设计中没有考虑到的一些意外因素和被试在实验中的体验，这些资料有利于实验结果的理解和解释。

第五节　数据的处理与分析

实验实施之后，研究者通常会面对大量数据，这些数据是在研究过程中记录的分数、测量值或观测值，接下来要完成的则是用统计学方法分析这些数据，使其隐含的意义显现出来。数据的统计分析主要有两方面的作用：第一，帮助研究者组织和概括资料，以便能更好地理解研究中发生的现象，并将研究结果与他人交流；第二，帮助研究者回答研究开始时提出的一般问题，确定研究结果所能证

① 林崇德、杨治良、黄希庭：《心理学大辞典》，上海教育出版社 2003 年版，第 523 页。

实的假设并做出结论。这两个基本的用途与两种基本的统计方法相对应:描述性统计(descriptive statistics)和推断性统计(inferential statistics)。

描述性统计是对一批数据的特点进行描述,它一般是将一批数据组织成图表的形式,并计算出一个或多个值,如用于描述数据整体情况的平均数(mean)、方差(variance)。描述性统计的作用就是组织、概括和简化一批数据。另一种方法,推断性统计则是利用来自样本的有限信息去回答有关总体的一般性问题。虽然样本来自总体,并且可以作为总体的代表,但并不能确保每一个样本都会准确地反映总体的情况,所以研究者必须慎重考虑,样本得到的结果能否反映总体的普遍性。推断性统计就是帮助研究者决定何种情况下能够用样本去反映总体。

实验结束后,对记录的原始数据及相关资料进行整理与分析,这主要包括初步整理、统计分析、结果的表述等三方面工作。

一、实验数据与资料的初步整理

这种整理主要是对实验记录进行汇总、归并、建档,其中包括数据资料,也包括一些文字记录资料。在一些较为复杂或大型的实验研究中,资料量是很大的,保证研究资料的清晰、有序非常重要,实验研究者要有一些有效的数据记录和整理经验。但在这一整理的过程中,一定要保证研究资料的原始性!

在前述整理的基础上,计算每一被试在某一特定实验条件下简单重复实验的平均值。比如在声光刺激简单反应时间比较的实验中,每一被试在声音刺激下或灯光刺激下重复测量反应时间 20 次,那就要将这 20 次测试结果平均以得到被试在声音刺激条件下的简单反应时间或灯光刺激条件下的简单反应时间。

最后,要形成一个由每一种实验处理条件下对应的每一被试测试结果的平均值组成的数据表。此表是后续统计分析的一个基本数据资料。

二、对数据进行统计分析

对数据进行统计分析的目的包括:对各种实验条件下被试反应总体情况的描述;对各种条件下被试反应之间差异性、相关性的分析和检验。就心理学的实验研究来说,前者经常用到的主要是计算一组数据的平均数(mean)、中位数(median)、标准差(standard deviation)、方差(variance)、四分位距、频数分布状

态，有时还需要对其分布的正态性进行检验和进行一些分数的转换，如标准分数转换等；后者经常用到的是两组数据间平均数差异的 t 检验（也有研究者常用 Z 检验）、多组数据间平均数差异的 F 检验、百分数差异的检验、计数变量的 χ^2 检验等。

上述数据分析的基本原理和过程详见有关附录 1 或参阅相关的统计学书籍。目前，心理学研究中数据的统计分析一般都借助统计软件来完成，如 SPSS、SAS 软件都很通用。

三、实验结果的表达

实验结果的表达主要包括三种方式：图、表和文字。其中图表的表达往往是必需的，这两种方式能非常直观地表现大量信息，而且这些图表具有通用形式，是同领域研究者都非常熟悉和认可的表达形式。从心理学的学术期刊可以看出，经常被用来反映心理实验结果的图形主要包括：线图（line graph）、条形图（bar graph）、直方图（histogram）、箱图（boxplot）等。在使用图形表达实验结果时，需注意，在图形的下边一般要有编号和标题，如“图 3 不同药物计量水平下血压降幅的比较”。

在心理学学术期刊上经常出现的数据表格形式，则主要包括描述性数据表、t 检验表、方差分析表等。关于数据表格的形式，这里强调两点：

第一，表格必须要有表头。表头由表的编号和标题组成，如“表 2 字频与语义启动对汉字语音识别影响的方差分析表”。

第二，表格尽量使用“三线表”。科技书刊中普遍采用三线表，即通常只有三条线：顶线、底线和栏目线。其中顶线和底线为粗线，栏目线为细线。当然，三线表不是说一定只有三条线，必要时可以加辅助线，但加了辅助线后仍称为“三线表”。

图表的信息要清晰、易读，表达方式要标准化，即采用科技文献，特别是心理学文献中通用的方式和符号。除使用图表以外，还要配以简明的文字表述。

第六节　研究报告的撰写

研究的最后阶段，是要将研究的过程与结果在科学界公开，目的至少有三个

方面:一是展示研究的成果,体现研究者的社会价值;二是接受同行专家的评估与监督;三是成果共享。为了这三个目的,研究报告必须实事求是、合乎规范,为此,中国心理学会(2002)专门组织编写了《心理学论文写作规范》。研究报告通常需要提供关于实验研究的以下三方面信息。

第一,你做了什么。报告要写明整个研究实施的过程,这一过程要描述得相当细致。

第二,你发现了什么。报告要对研究结果进行客观介绍和评估,这要写明采用的测量方法、统计分析手段以及对这些结果作何种解释等。

第三,你的研究与相关领域中的其他研究有什么联系。一项质量较高的实验研究往往不是孤立的,而是缘起于已有的相关知识体系和在此基础上的进一步发展。研究报告应该体现当前研究和过去研究的发展关系。

研究报告是非常结构化的文件,它由若干部分组成,每一部分的内容都有特殊的、明确的规定,你需要根据每一部分的要求按部就班地介绍你自己的研究过程和结果。下面我们按照顺序介绍研究报告每一部分的具体内容和写作要求。

一、标题、作者和机构

标题(title)是对文章内容的精练表述,它应当明确地表达被考察的变量、理论及关系,尽可能完整准确地概括出研究的内容,又要避免使用多余词语。你要特别记住,论文标题中使用的语词将成为他人查阅和检索你论文的关键线索,而且人们往往从对论文标题的第一印象来决定是否还要继续阅读这篇文章的其余部分。论文题目长度应尽量控制在 20 个汉字以内。

紧挨着标题的是作者(author)姓名,然后是作者供职的机构或研究实施的机构,并附带给出其所在的城市和邮编,如"(南京师范大学心理学系,南京 210097)"。当有多名作者,排序通常是重要的,排在第一位的作者一般是对研究作出主要贡献的人,然后按贡献大小依次排列其他作者的姓名。

二、摘要和关键词

摘要(abstract)要用非常简练的语句表述研究的目的、研究的手段和方法、研究的主要结果或发现。摘要通常都是在论文的其他部分全部写完以后才写

的，它是写作中应特别重视的部分。除标题外，摘要是研究者在搜索文献时首先阅读的部分，而且他们会根据摘要提供的内容来决定是否进一步查找和阅读全文。

一篇实证研究报告的摘要不是论文写作的引导段落、补充说明或评价，而是具有相对的独立性，其长度应控制在200个汉字或500个英语单词以内。它一般由以下各项内容组成：对研究问题或目的的一句话概述；对研究被试的简单说明（要明确说明被试数量和相关特征）；对研究方法和步骤的简单描述；主要结果的报告；结论或意义的精练陈述。这些内容不一定按照固定顺序来写，也未必面面俱到，但通常都要写到：研究问题或目的、研究方法或手段、研究结果和发现等内容。

在行文方式上，要尽量使用第三人称。比如采用“对……进行了研究”、“报告了……的现状”、“进行了……的调查”等记述方式，不必使用“本文”、“作者”等作为主语。① 另外，需提醒的是，不少初写论文者常常会把摘要写成“引言”或“前言”的性质，这一点要避免。

摘要之后，给出3～5个关键词(keyword)。关键词是对描述研究范围或领域、研究核心内容、采用的方法等起重要作用的词汇，或者其在文中出现的频率较高。关键词在文献搜索中的作用很大，因此撰写研究报告时，要重视关键词的选用。

三、引言

研究报告主体或正文的第一部分是“引言”(introduction)，也叫做“序言”、“前言”或“问题提出”、“研究目的”等。“引言”主要是向读者介绍研究的背景和方向，以及提出此项研究的理由。具体要说明：你的研究要解决什么样的问题或疑惑？是如何想到要解决这一问题或疑惑的？这样做有什么意义或重要性？这一研究的提出和结果与该领域现有的知识有什么样的联系？等等。通常作者还会在此部分的最后以假设的形式明确实验的目的。

在初写研究报告时，不妨按照下列程式来写“引言”，熟练之后再灵活地体现

① American Psychological Association(1995). *Publication manual of the American Psychological Association* (4th ed.). Washington, D.C.

出自己的风格：

第一，引题。用较少文字说明研究什么问题，以及研究的必要性、重要性或意义。

第二，文献回顾。文献回顾的目的在于说明你提出研究问题的背景、理由和论据，一般的逻辑路线是：在相关领域中，前人做了哪些主要的研究，这些研究给予我们什么样的认识，这些研究还有哪些空白区，还有哪些疑惑或相互矛盾的地方，这里的某一空白区或矛盾需要解决，为此需要进行什么样的研究，实际上就是将读者的思路带入研究主题。文献回顾切忌两点：一是不加选择地论及该领域的全部或大部分研究，二是不加整合地罗列文献。

第三，界定关键词或变量。文献回顾后，研究问题就越来越明晰了，这时对涉及的主要概念或变量进行界定，就可以将研究的问题限定在某一具体范围。

第四，研究思路或方法的介绍。简单介绍研究的方法论和基本思路，描绘出研究中准备使用的方法的大致轮廓，可以使读者思想上有所准备。

第五，提出明确的假设。在引言部分的最后，一般都要提出关于研究变量关系的假设，这既是对结果的预期，也是研究方向与目标的再一次重申。

四、方法

研究报告正文第二部分是研究的方法(method)。方法部分就研究实施过程进行相对详细的介绍，以使其他研究人员能够清楚你的研究方法，从中获得的信息足以使他能够复制你的研究过程，如果他想要这样做的话。方法部分一般包括四方面内容：研究被试、仪器设备与材料、研究设计类型、实验操作程序与主要步骤。

“被试”部分一般要说明：(1)被试数量；(2)如何抽样和分组的；(3)被试基本的人口统计学特征，包括年龄、性别、种族等；(4)与研究相关的其他特征(例如，IQ或精神病理学特征)。对于动物被试来说，提供的信息与人类被试类似，如：(1)使用的动物数；(2)它们的种群、种类和血统；(3)提供者；(4)这些动物是如何居住和饲养的；(5)具体特征，包括性别、重量、年龄等。

“仪器设备与材料”(apparatus and materials)一般要说明研究中使用的器材(如：设备)或材料(如：调查问卷、实验用词单等)。有时这两个部分同时包括在研究报告中。在器材部分，普通的项目如椅子、桌子、秒表等不需要介绍得太详

细，越是特殊的器材，越是要给予详细说明。对于定制的设备，还要求有图形或图片。如果器材只需简单提及，有时干脆忽略或与方法中的其他部分混在一起。对于使用调查问卷的研究，这一部分就叫做“材料”。在研究中用到的每份调查问卷都要有相应的介绍，并且要对它的作用进行说明（例如：用什么来进行测量）。对于那些新编制的调查问卷，还必须将其放在附录里。

“研究设计类型”，主要说明采用了什么样的实验设计模式，如“研究采用2×3×2混合实验设计”、“研究采用2×3完全随机区组设计”。

“实验操作程序与主要步骤”要对研究中变量的操控方法与研究实施步骤作较为详细的说明，包括被试分组方法、给予被试的指示语、较特殊的测评步骤等。对于实验研究来说，还要详细说明为了避免变量混淆而进行的操作和控制措施。

五、结果与分析

“结果与分析”（results and analysis）部分要对研究中得到的资料进行概括，并进行统计分析。结果的展示顺序通常是：陈述主要的研究成果，给出基本的描述性统计结果（如平均数和标准差），再给出推断性分析结果（一般是指假设检验的结果），最后给出效应量的测量结果。

结果的呈现最好使用图表的形式。图表对信息的呈现直观、易读、易懂，再辅以简单的文字说明就更清楚了。当然，如果研究本身很简单，能很容易地用文字说明，也可以不使用图表。在“结果与分析”部分，要注意两点：一是客观说明得到的结果，这里的“分析”不是“讨论”之意，而是“统计分析”之意，在此部分“切忌夹叙夹议”，[①]防止将研究者的主观认识与研究中得到的客观结果混淆在一起；二是结果的呈现不能随意罗列，要结合前文提出的研究目标和研究假设的顺序与层次，进行结构上的组织和编排。

要报告统计学意义上的显著性，其陈述的内容要给出以下信息：(1)使用的检验类型；(2)自由度；(3)检验结果；(4)显著性水平。当报告显著性水平时，鼓励使用精确的概率值（其大多是由计算机程序提供的），或者使用传统的 α 水平

① 有幸师从黄希庭教授，其间的1996年初，不无得意地将一篇自以为还不错的研究报告交到先生手中，几天后先生约见。先生对我的报告做了详细审阅与批注，迄今印象最深的就是在“结果与分析”部分的页边上的几个大字：“切忌夹叙夹议。”后来，我也就学着先生的口气告诉我的学生：切忌夹叙夹议！

(如 $p<0.05$、$p<0.01$、$p<0.001$ 等)作为参考点。例如:

使用精确概率,可以这样报告结果:

结果显示,两组之间的差异显著,$F(2, 36) = 4.37$, $p = 0.006$。

使用传统的 α 水平,则可以这样报告结果:

结果显示,两组之间的差异显著,$F(2, 36) = 4.37$, $p < 0.01$。

六、讨论

研究报告正文的第四部分是“讨论”(discussion)部分。在讨论部分,你要对研究发现的意义进行解释、评价和讨论。“讨论”应从假设的重述开始(回想一下你最初在引言部分结束前提出的假设),接着简要重述你的主要结果,指出它们是否支持你的假设,然后将你的结果与其他研究者的结果联系起来,说明你的结果在多大程度上符合该领域现有知识结构。通常,还要辨析研究的所有缺陷,特别是影响结果普遍性的因素。

初写研究报告者,往往不知如何展开讨论,这时可将结果与前文提出的研究假设、与该领域其他研究进行比较,这会帮助你意识到哪些问题需要解释:(1)研究结果验证假设了吗?如果验证了假设则意味着什么?如果只是部分验证或全部未能验证则意味着什么?为什么会如此?(2)研究结果与该领域有关理论和相关研究吻合吗?吻合意味着什么?不吻合意味着什么?为什么不吻合?(3)研究存在什么问题?为什么未加以控制或克服?这可能带来什么样的后果?(4)研究有无引出新的矛盾?随后需要进一步探讨的问题是什么?

讨论并不复杂,它就是对“结果”中隐含的意义作进一步的发掘和说明,帮助阅读者理解实验研究的结果,也是在为“结论”做准备。在讨论部分,要注意回应“引言”,切忌抛开研究假设和目的,不着要领、离题太远。

七、结论(与建议)

研究报告正文的最后部分是“结论”(conclusion),有时叫做“结论与建议”。这一部分就是将本研究中确实得到的、可以肯定的部分概括性地呈现出来,它是前述“结果与分析”、“讨论”后的一个自然总结。这里也需要注意两点:一是不能将前人已有的认识和他人研究的结果作为“结论”,二是不能将研究者的假想、未

能肯定的推测、未来研究设想等作为“结论”。结论要简短、肯定。

在一些应用性研究的报告中，研究者往往根据研究的结论针对某一实践领域提出建议。一般，这些建议比较简明、确定，不要作过多的阐释，以防止有“喧宾夺主”之嫌。

八、参考文献和附录

“参考文献”(reference)和“附录”(appendix)已不属于研究报告的正文部分，但同样是重要的。参考文献至少有四方面的作用：(1)为自己的研究及讨论提供支持，并证明研究者本人对该研究领域的了解程度；(2)向被引用文献的原作者表示谢意，并给予应有的声誉；(3)承认引用，避免“抄袭”、“剽窃”之嫌；(4)为阅读者提供进一步查阅相关文献的线索。所以，“参考文献”要为报告中每一引用项提供完整信息，而且要注意，参考文献里列举的项和论文里引用的项必须一一对应，即每个引用项都必须出现在参考文献列表里，而参考文献列表里的每一项又必须是被引用的。一般，参考文献按第一作者姓氏的字母顺序排列。第一作者姓氏字母相同时，独著类的先列，其他则依照时间先后顺序列。也有的是按照引用的先后顺序排列。

不过，就国内学术刊物来说，参考文献的编写方式各有所不同。投稿前需要查阅相关刊物最新的版式和最新的“征稿启事”。

“附录”部分一般是对有关的仪器设备、实验材料作详细说明或列举，有时研究者也会把实验的相关数据信息列举于此。

阅读材料 1-1

心理学研究中的主要资源

常用的资料文献主要有专著、专业期刊、学位论文、学术会议论文集、索引、电子出版物、电子学术期刊和网络文献数据库等。

1. 专著

专著通常汇集了某一学科领域或具体问题的研究成果，因此，通过查阅某一

领域或具体研究问题方面的专著，可以使我们对该领域或者研究问题有较为系统和全面的了解及掌握，便于选择新的研究课题和开展深入的研究。

2. 专业期刊

这是研究者查阅的主要文献资源。在心理学研究领域，国外核心期刊主要是收录在SCI(Science Citation Index)和SSCI(Social Science Citation Index)上的杂志，国内的心理学核心期刊主要收录在CSCI和CSSCI上的杂志(如:《心理学报》、《心理科学》、《心理科学进展》、《心理学探新》、《心理发展与教育》等)。通过查阅核心专业期刊可以获得某一学科的不同专业领域和研究方向研究的历史和现状，以及研究者普遍关注的热点问题。

3. 学位论文

包括硕士学位论文和博士学位论文，这类论文主要集中在高等院校和科研机构，并收录在高等院校或研究机构的图书馆中。近年来，学位论文也纳入到国家图书馆收藏的文献之列，在国家图书馆设有博士学位论文查阅专库(该库中也有部分优秀的硕士学位论文)。读者可以查阅、复制。学位论文的研究内容具有前沿性，对问题的研究和阐述较为详细，且相当部分的学位论文没有发表或部分没有发表，所以学位论文具有较高的参考价值。

4. 学术会议论文集

一些重要的国际性、全国性学术会议都会出版论文集。这样的论文集往往是选择一次会议中提交的具有代表性的论文，可以反映一个国家或国际范围内心理学的最新发展状况。特别是一些专业性很强的学术会议，其论文集基本上反映了该领域国内外的最新进展。这些论文集一般也可以从国家图书馆中查到。

5. 索引

索引是传统的文献查询途径。现在有了计算机技术的支持，可以将索引制作成光盘或在线数据库，使得查询非常便利、快捷。目前，在心理学领域内，信息量最大的索引或光盘数据库当数美国心理学会发表的《心理学文摘》(Psychological Abstract)光盘数据库。可到国家图书馆电子阅览室查询或某些大学图书馆查询。

6. 电子出版物、电子学术期刊和网络文献数据库

随着各高校办学条件的改善，各高校图书馆引入了越来越多的电子出版物、电子学术期刊和网络文献数据库。与心理学关系密切的如：中国学术期刊全文

数据库、Elsevier、PreQuest、EBSCO、PsycARTICLES、PsycINFO、Psychology and Behavioral Science Collection 等。

——资料来源:《实验心理学纲要》(张学民、舒华,2004)①

阅读材料 1-2

出版的伦理道德

出版的伦理道德是一个原则性问题……道德原则的遵守是为了两个目标:(1)保证科学知识的准确性;(2)保护知识产权。出版伦理道德涉及以下方面的内容:

1. 实事求是地报告研究结果

科学方法的本质是能够被他人重复和验证,所以应该如实地报告研究过程和研究结果,不应伪造数据或篡改结果以迎合原先提出的理论假设,也不应忽略研究中的错误。如实际被试多少人,有效被试多少人;实际观测数据有多少,参与统计处理的有效数据有多少,哪些极端数据被删除,为什么要删除,这些都应在报告中如实说明。研究者不应为了所谓研究的"漂亮"而忽略一些无效的数据,更不应该删除那些不利于研究假设但却是有效的数据。如果做了 5 个实验,只有 3 个实验比较理想,虽然研究结果只包含了 3 个实验的结果,但在研究报告中也要说明是做了 5 个实验而不是 3 个实验。

2. 剽窃

在论文写作中,如果直接引用其他研究者的原话,应该把原话用引号括起来,并注明出处;对其他研究者的思想的重新阐释,也应注明出处。关键是不要把别人的思想或工作当作自己的。尤其是在研究报告的问题提出和讨论部分以及综述性文章中,当引用或参考他人的思想或工作时,一定要注明出处,在文后的参考文献表中列出来。这不仅是对他人劳动的尊重,也对心理学的健康发展起到非常重要的作用。如果所有的引文都不注明出处,势必会给整个研究领域带来混乱,读者(包括作者自己)也就无法知道某个研究思想的源头在哪里。

① 张学民、舒华:《实验心理学纲要》,北京师范大学出版社 2004 年版,第 16—18 页。

在研究中引用他人的资料时,应注意以下几点:(1)直接引用少于40个字用引号;(2)多于40个字应另起一段;(3)引用要忠实原文,即使原文是错误的;(4)省略部分用省略号;(5)额外的解释放在括号内;(6)不要忽略引用中的引用,但在参考文献中不必列出。

3. 论文的署名

论文作者不仅是执笔人,还包括那些对研究做出过实质性科学贡献的人,他们可能参与提出问题假设、构思实验设计、统计分析、结果解释以及文章主要部分的写作。对于辅助人员……应在正文后的致谢栏里注明……

论文署名的基本原则是按贡献的大小排序。在有多个研究者参与的合作研究项目中,合作者之间应该事先商定好研究的分工、谁有论文署名权以及署名的先后。在论文中所有署名的作者都应对文章的准确性负责,所以,在论文投递前,每个作者都应该认真审阅。一旦论文被接受,有的期刊还要求每个作者都应签名确认。

4. 论文的重复发表

……论文的重复发表还会导致版权纠纷……不应为提高命中率而一稿多投……

5. 原始数据的保存

……在文章发表之后,作者还应把原始数据保存5年以上,以备他人查证或重复分析。

——资料来源:《心理学论文写作规范》(中国心理学会,2001)①

建议阅读文献

1. 杨治良编著:《实验心理学》,浙江教育出版社1998年版,第4—29页。
2. 黄希庭:《心理学实验指导》,人民教育出版社1988年版,第1—39页。
3. 张学民、舒华编著:《实验心理学纲要》,北京师范大学出版社2004年版,第14—23页。
4. 中国心理学会编:《心理学论文写作规范》,科学出版社2001年版。

① 中国心理学会编:《心理学论文写作规范》,科学出版社2001年版,第88—91页。

复习思考题

1. 如何理解变量、刺激变量、机体变量、反应变量、自变量、因变量、额外变量、信度、效度、区分度、样本、代表性样本、有偏样本、心理实验设计、组间设计、组内设计、指示语、指示语效应、实验者效应(罗森塔尔效应)、霍桑效应?
2. 心理学实验研究课题的来源有哪些方面?
3. 如何从文献检索与分析中寻找研究的切入点?
4. 如何从已有理论的分析中寻找研究的切入点?
5. 文献检索与分析的作用有哪些?
6. 心理实验研究中的变量有哪些? 分别如何处理?
7. 心理实验中因变量的种类和条件有哪些?
8. 心理实验中常用的额外变量的控制方法有哪些?
9. 实验被试的抽样方法主要有哪些? 各有什么优缺点?
10. 什么叫做组间设计、组内设计、混合设计? 分析它们的优缺点。
11. 制订实验指示语需要注意哪些问题?
12. 心理实验中主试与被试的相互作用表现在哪些方面?
13. 心理学实验报告主要包括哪些部分? 各部分的主要内容是什么?
14. 列出若干你想到的或经过阅读相关文献后提出的研究课题,并对其中你觉得最有意义的一两个课题进行分析,阐释其意义和研究实施可能需要的条件。

第二章
多因素心理实验设计

本章内容提要

心理实验的核心本质在于验证自变量与因变量间的因果关系，这里的自变量叫做因素或因子，因此心理实验也叫做因素型实验。根据实验中被操纵自变量的数量，因素型实验设计包括单因素实验设计和多因素实验设计，其中多因素实验设计又叫析因实验设计。本章对因素型实验设计的类型进行简单分析之后，主要介绍析因实验设计的基本原理和数据分析方法，包括多因素完全随机实验设计、多因素重复实验设计、混合实验设计，最后将研究设计扩展至区组设计和拉丁方设计。

第一节　因素型实验设计的类型

心理学的实验研究主要是一个控制额外变量、操纵自变量、观测记录因变量、验证自变量与因变量之间因果关系(causality)的过程，因此心理学实验也叫做因素型心理实验，简称因素型实验(factorial experiment)。其中自变量也叫做因素或因子(factor)。对自变量的操纵(manipulation)是心理学实验的主要特征之一。

在因素型心理实验中，无论怎样控制额外变量和操纵自变量，归根结底，还是要对被试的心理或行为变化进行观测、记录，所以它实际上就是一种有控制的观察。从这一角度看，观察法和实验法具有共通性，它们可被看作同一维度上的两个不同区，就如图 2-1 所示的那样。观察法是在保证被试完全自然存在的前提下，对其心理和行为进行不加干预的直接观察，据此描述其心理活动规律的方法。实验室实验法则是在严格控制的实验室条件下，对被试的心理和行为进行观察，据此推断其心理活动规律的方法。准实验法和自然实验法

均为介于观察法和实验室实验法之间的心理学研究方法，这两种方法都是指对被试有一定的干预和影响，但对实验中可能的额外变量未作严格控制的研究方法，其在一定程度上保证了研究对象的自然存在性。在图 2-1 中的坐标轴上，越靠近右端，实验中对变量的控制或操纵越严格；越靠近左端，实验中对变量的控制或操纵越少。

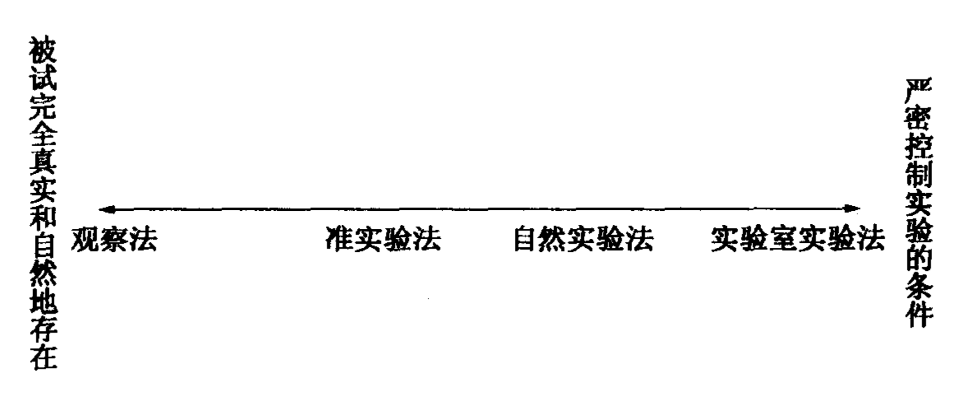


图 2-1　同一维度上的研究方法的比较

本书不讨论观察法，将讨论的范围限定在准实验设计和真实验设计。真实验设计就是要在实验过程中尽量严密地控制实验条件，以探求被试心理活动的因果关系。准实验设计则对研究条件的控制不甚严密，特别是研究中的被试组，往往不是经过随机化或匹配方法得到的相等组，研究中很可能存在成为混淆因子的机体变量或其他变量。自然实验法实验，是在操纵和控制了一系列变量的同时，尽量使研究情境具有自然性和真实性。因此，准实验法、自然实验法、实验室实验法三者间的区别是相对的。

实验设计和准实验设计(quasi-experimental design)存在明显不同，主要是被试组间的相等性(自变量之外的变量的相等性)问题。实验设计的被试组一般有两种组建方式：第一种组建方式是，从同一总体抽取被试，然后按照随机化方法或匹配法将被试分配到不同的实验处理中，形成相等组；第二种组建方式是，根据个体在某一机体特征上的不同而将研究对象划分成子群，然后从每一子群抽取一个被试组，这些被试组除了在分组变量及相关特征的水平上不同外，其他特征基本一致，这里的分组变量是一个准自变量。准实验设计的被试组也有两种组建方式：第一种方式是，由于受某些条件的限制，研究者虽能组建被试组，但却无法做到被试组相等，比如用一组多动症儿童进行相关的药物试验，同时组建一个同年龄的正常儿童组作为控制组，这两个组就不可能是等组的，进行的研究就属于准实验设计；第二种组建方式是，直接选用现成群

体作为被试组，如不同班级的学生、不同工作车间的工作小组、不同医院里收诊的某类病人等，这些被试组一般都很难是相等的组。即使只使用一个被试组，准实验设计也难以平衡某些额外变量，所以准实验设计中往往存在明显的、未加控制的混淆变量。概括地说，准实验设计中的混淆变量主要是被试组间的差异性、实验过程中新增的额外变量，这些因素使其结论具有一定程度的不确定性，即其内部效度相对较低。但由于准实验设计的研究情境较接近人们真实的心理生活实际，如果能有效把握实验进程，其得到的结果可能会具有较高的外部效度。

现在考虑另外一个问题：心理学实证研究的目的是什么？归纳来说，其目的大都属于三方面：第一是为了测定人的心理特征，如测定感受性、视敏度、记忆力、智商水平、气质特征等，这类研究可以用实验方法进行测试，可以用观察法进行鉴定，还可以使用心理测验量表进行测定，因此这类研究可统称为测验式实验；第二是为了初步探测某一心理现象的可能影响因素或它对其他心理现象可能存在的影响效应，采用大样本的多维度测验，然后进行相关分析，这一类研究可以称之为相关实验研究或相关研究；第三是为了直接探明某种心理现象与其他心理现象或内外因素的因果关系，这叫做因素型实验或函数型实验。我们准备重点讨论的实验设计主要从因素型实验的角度考虑。因果关系的研究可以是观察法、准实验法和实验法。观察法不在本课程的讨论范围，我们就讨论因素型实验设计和因素型准实验设计。

那么，因素型实验设计有哪些类型？这要从以下三个维度来分析。

第一，自变量数。分为一个自变量的和两个或两个以上自变量的两大类，即单因素实验设计(one-factor experimental design)、多因素实验设计(multi-factor experimental design)，多因素实验设计又可以分为二因素实验设计(two-factor experimental design)、三因素实验设计(three-factor experimental design)……比如，要研究选择反应时间是否受到灯光刺激颜色的影响，自变量就只有一个，即刺激光的颜色，叫做单因素实验设计；要研究左右耳对于不同频率的声音刺激的感受性是否存在差异，则自变量有两个，一个是听觉通道(有左耳和右耳两个水平)，另一个是声频(声频可以根据研究者对研究精确性、条件许可程度设置两个或两个以上的水平，比如以低音、中音、高音三个档次，或以 500 Hz、2 000 Hz、10 000 Hz 等)，这一实验设计叫做多因素实验设计或二因素实验设计。

第二，被试的选择和分组方法。被试选取和分组的方法可以是随机的，

即从研究对象的总体中随机抽取一定量的被试数，在分组时也是按照随机的方法；也可以采用匹配法，即对被试先进行某些心理品质的预测，根据预测的结果把某种心理品质接近的被试分配到不同的实验处理中，以使各实验组被试基本相等。当然，还可以使用一个被试组进行实验，那就不需要对被试进行分组。因此，按照第二个维度，就可以将实验设计分为：独立组设计（采用随机方法对被试分组）、配对组设计（采用匹配法对被试分组）、单组设计（不对被试分组）。

第三，实验处理的分派。这里有三种情况：第一种情况是，每一被试只完成一种实验处理，这就构成了独立组设计或配对组设计；第二种情况是，一个被试组承担所有的实验处理，这就构成了重复实验设计；第三种情况是，每组被试完成某一个或多个自变量所有水平的实验处理，但只完成另一个或另几个自变量的一个水平的实验处理，这就构成了混合实验设计(mixed experimental design)。

按照上述三个维度将主要的实验设计类型表示成图 2-2 所示的形式。在图 2-2 中，我们还列出了区组实验设计和拉丁方实验设计，这是在上述基本类型的基础上扩展出来的较为特殊的研究设计，本书只介绍在完全随机实验设计基础上扩展出来的区组实验设计和拉丁方实验设计。

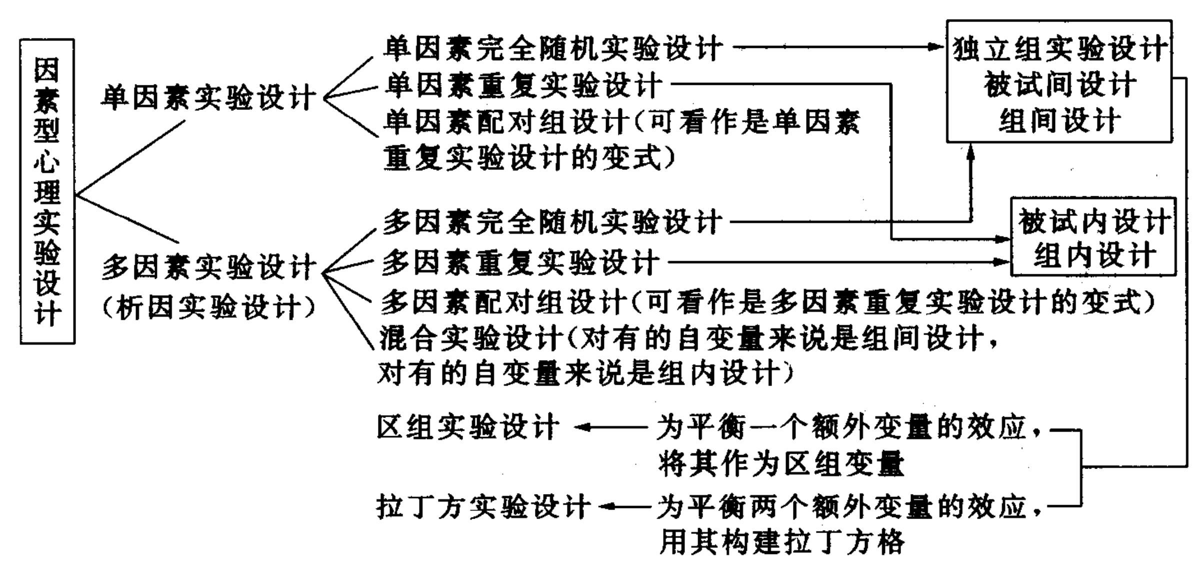


图 2-2 心理实验设计类型简图

按照哪种方法来设计实验，需根据实验的具体课题和实验的条件，特别是变量的性质来选择。在本章中，我们拟对多因素完全随机实验设计、多因素重复实验设计、多因素混合实验设计、区组实验设计和拉丁方实验设计进行介绍。

第二节　典型的多因素实验设计

一、多因素完全随机实验设计

对于单因素完全随机实验设计来说,实验处理数就是自变量的水平数,将被试随机分配到各实验处理单元就可以了。多因素完全随机实验设计则是多个因素多种水平相结合,构成多个实验处理,其实验处理数等于所有自变量的水平数之积。如二因素二水平实验,就是有两个自变量,每个自变量有两个水平,则结合形成的实验处理数就是 2×2＝4,这种实验设计被称为2×2 完全随机实验设计;如果一个自变量有两个水平,另一个自变量有三个水平,则共有 2×3 ＝ 6个实验处理,这种实验设计被称为2×3 实验设计。如果有三个自变量,其中两个自变量各有 2 个水平,第三个自变量有 3 个水平,则这种实验设计有 2×2×3 ＝12 个实验处理,就叫做 2×2×3 实验设计。这里需要重申以下几点:

第一,自变量是研究者操纵的变量,在实验过程中必须是变化了的,也就是说自变量的水平数至少为 2。如果自变量的水平数为 1,那就等于说该变量在实验过程中始终保持在一个水平上,它就不是"变"量了。比方说,一个 2×3×1×2 实验设计中,实际上只有三个自变量,它们的水平数分别为 2、3、2。

第二,实验处理数等于所有自变量水平数的乘积。如一个 2×3×3 实验设计,其实验处理数是 18,等于说这一实验过程中出现了 18 种实验条件或实验处理。

第三,对于完全随机实验设计来说,有多少种实验处理就要有多少组实验被试,因为一组被试只参加一种实验条件下的实验。

现在,我们以一个假想的实验研究为例来说明多因素完全随机实验设计(multi-factor randomized experimental design)的模式。

假设某研究者想考察缪勒-莱尔错觉(Müller-Lyer illusion)受箭头方向和箭头角度的影响。研究中的自变量有两个:一个是箭头方向(标记为 A),分为向外(A_1)和向内(A_2)两个水平;另一个是箭头角度(标记为 B),设置为 15°(B_1)和 45°(B_2)两个水平,因此这是一个 2×2 实验设计,构成了 4 种实验处理,如表 2-1 所示。研究者从某大学文学院本科二年级学生中随机抽取了 20 名男生,再将这 20 名男生随机分成相等的四组,每组 5 人,每一被试组接受一种实验处理。因此,这是一个二因素完全随机实验设计。假设实验得到表 2-1 的数据,那么如何

分析这些数据呢?

表 2-1 箭头方向与箭头角度对错觉量的影响

	箭头方向向外(A_1)		箭头方向向内(A_2)	
	箭头角 15°(B_1)	箭头角 45°(B_2)	箭头角 15°(B_1)	箭头角 45°(B_2)
	6	4	8	7
	5	3	7	6
	7	5	9	7
	6	4	8	6
	7	5	9	8
$\sum$	31	21	41	34

这一数据分析的目的就是要考察自变量的变化是否引起了因变量的变化。具体地说,就是箭头方向的改变是否导致了缪勒-莱尔错觉量的不同、箭头角度的改变是否导致了缪勒-莱尔错觉量的不同、这两个自变量对因变量的影响是相互独立的还是相互依赖的呢?根据统计学方法,拟采用完全随机实验设计的方差分析(analysis of variance)来确定是否存在上述效应。这一方差分析的过程如下:

第一步:计算数据总变异量并对之进行分解

显然,表 2-1 中的 20 个数据之间是有大小差异的,即数据存在变异,变异的原因大致可以划分为四个方面:(1)自变量 A 的独立作用,叫做 A 的主效应(main effect);(2)自变量 B 的独立作用,叫做 B 的主效应;(3)自变量 A 和自变量 B 的交互作用,叫做 A 和 B 的交互效应 ($A \times B$ interaction)①;(4)来自被试间差异及其他随机因素的影响,我们将之称为误差效应,或残差(residual error)②。

按照统计学方法,一组数据的变异量是用该组数据的离差平方和来计算的,所以也叫平方和(sum of square,简称 SS)。平方和的通用计算公式是:

$$SS = \sum (X - \overline{X})^2 = \sum X^2 - (\sum X)^2 / N$$

就本例来说,当要计算表 2-1 中 20 个数据的总变异量时,就是直接计算表中 20 个数据的离差平方和:

① 所谓交互效应,就是不同的变量必须同时存在并发生变化时才能产生的对因变量的影响。

② 毫无疑问,残差部分包含了多方面因素的共同作用,但由于这些因素的变化未将数据分组,所以难以单独估算,就笼统地称之为残差。残差越大,说明观测数据的变化是由偶然因素随机变化引起的可能性就越大;反之,数据变化由于操纵变量引起的可能性就越大,这就是方差分析的逻辑。

$$SS_T = \sum(X-\overline{X})^2 = \sum X^2 - (\sum X)^2/N$$
$$= \sum\sum X_{ij}^2 - (\sum\sum X_{ij})^2/N$$
$$= (6^2+5^2+\cdots+6^2+8^2) - 127^2/20$$
$$= 859 - 806.45 = 52.55$$

当要计算自变量 A 单独引起的数据变异量时，其计算逻辑是：当自变量 A 取两个水平 A_1 和 A_2 时，表 2-1 中的数据就被 A_1 和 A_2 分成了两个小组，每个小组 10 个数据。如果我们以每个小组的平均数替代该小组中的每个数，这样构成的 20 个数据之间的变异量就全部是由自变量 A 的变化引起的，其他因素引起的数据变异在平均过程中被平衡了。也就是说，此时计算这被替换后的 20 个数据的离差平方和，就正好是自变量 A 单独作用引起的变异平方和：

$$SS_A = \sum(X-\overline{X})^2 = \sum X^2 - (\sum X)^2/N$$
$$= \sum_j \left[\frac{\sum_i X_{ij}}{10}\right]^2 \times 10 - (\sum\sum X_{ij})^2/N$$
$$= \sum_j \frac{(\sum_i X_{ij})^2}{10} - (\sum\sum X_{ij})^2/N$$
$$= [(31+21)^2/10 + (41+34)^2/10] - 806.45 = 26.45$$

同样道理，计算自变量 B 单独作用引起的变异平方和：

$$SS_B = \sum(X-\overline{X})^2 = \sum X^2 - (\sum X)^2/N$$
$$= [(31+41)^2/10 + (21+34)^2/10] - 806.45 = 14.45$$

如果自变量 A 和自变量 B 同时发生变化，则数据被分成了四个小组。按照上述同样的道理，可以计算 A 和 B 同时变化时引起的变异平方和。这一变异平方和包含了自变量 A 和自变量 B 各自单独作用引起的变异平方和，以及二者交互作用引起的变异平方和，所以要计算 A 和 B 交互作用引起的变异平方和时，其计算如下：

$$SS_{AB} = \sum(X-\overline{X})^2 - SS_A - SS_B = \sum X^2 - (\sum X)^2/N - SS_A - SS_B$$
$$= \sum_j \frac{(\sum_i X_{ij})^2}{5} - (\sum\sum X_{ij})^2/N - SS_A - SS_B$$

$$= (31^2/5+21^2/5+41^2/5+34^2/5)-806.45-SS_A-SS_B=0.45$$

最后,可以计算残差项的平方和:

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$
$$= 52.55 - 26.45 - 14.45 - 0.45 = 11.2$$

第二步:计算各因素引起变异量对应的自由度(即数据发生变异的次数)

因观测得到 20 名被试的数据,所以总变异对应的自由度 $df_T = 20-1=19$。然后对其分解:

A 的主效应对应的自由度: $df_A = a-1=1$ (a 是自变量 A 的水平数)

B 的主效应对应的自由度: $df_B = b-1=1$ (b 是自变量 B 的水平数)

A 和 B 的交互效应对应的自由度: $df_{AB} = (a-1)(b-1)=1$

残差的自由度等于总自由度减去上述三项自由度: $df_E = 19-3=16$

第三步:计算各变异源引起数据变异的方差,即均方 *MS*

因为方差等于变异平方和除以自由度,于是:

$MS_A = SS_A/df_A = 26.45$

$MS_B = SS_B/df_B = 14.45$

$MS_{AB} = SS_{AB}/df_{AB} = 0.45$

$MS_E = SS_E/df_E = 11.2/16 = 0.7$

第四步:计算各效应是否显著的检验统计量 *F* 比率

也就是计算各效应方差与残差方差的比值:

$F_A = MS_A/MS_E = 26.45/0.7 = 37.786$　分子与分母的自由度为(1, 16)

$F_B = MS_B/MS_E = 14.45/0.7 = 20.643$　分子与分母的自由度为(1, 16)

$F_{AB} = MS_{AB}/MS_E = 0.45/0.7 = 0.643$　分子与分母的自由度为(1, 16)

第五步:给出方差分析表和分析结论,如表 2-2 所示

查 F 表确定各效应 F 比率达到统计学上的显著性水平所需的临界值,得到:

$$F(1, 16)|_{\alpha=0.05} \approx 6.20 \quad F(1, 16)|_{\alpha=0.01} \approx 10.80$$

将上述计算得到的 F 比率与临界值比较,就可以确定各效应的 F 比率是否达到显著性水平。从比较的结果知道: F_A 和 F_B 均大于 $F(1, 16)|_{\alpha=0.01}$,但 F_{AB} 小于临界值 $F(1, 16)|_{\alpha=0.05}$ 和 $F(1, 16)|_{\alpha=0.01}$。将上述分析的结果汇总,得到方差分析表如表 2-2 所示。

表 2-2　箭头方向与张开角度对错觉量影响的方差分析表

变异来源	变异平方和	自由度	均方	F	p
A的主效应	26.45	1	26.45	37.786	<0.01
B的主效应	14.45	1	14.45	20.643	<0.01
AB的交互效应	0.45	1	0.45	0.543	>0.05
残　差	11.20	16	0.70		
合　计	52.55	19			

从方差分析表可以看出，自变量A和自变量B的主效应达到了显著性水平($p<0.01$)，A和B的交互效应没有达到显著性水平($p>0.05$)。因此，得到结论：箭头的方向和张开角度对缪勒-莱尔错觉量有显著性影响，且二者对错觉量的影响是相互独立的(***提醒：此结论由虚构的表 2-1 中数据得到，非实验研究结论***)。

虽然，这里所举例子是最简单的多因素完全随机设计，但它能够说明完全随机设计的几乎所有特征，包括如何评估自变量的主效应和交互效应。如果遇到自变量或自变量的水平数更多的实验设计，其实验的原理和数据分析的程序都与这里展示的过程相同。比如，对于一 2×3×2×4 完全随机实验设计来说，其自变量是 4 个，实验处理数是 48，那么实验就需要 48 组被试。在数据分析中，需要分析四个自变量的主效应、两两变量间的交互效应、三个变量间的交互效应、四个变量间的交互效应等，这里需要考察的交互效应就达 11 个。显然，随着自变量数和变量水平数的增加，特别是被试数的增加，会给方差分析带来非常繁琐的计算。不过，这一点不用担心，因为在实际研究中，研究者都是使用统计软件进行数据分析，一切都变得相当快捷了。

可还存在另一个问题，即随着变量数和变量水平数的增加，实验处理数急剧增加，这就意味着被试组数的大幅增加。对于具有 48 种实验处理的实验来说，需要 48 组被试，如果再考虑每一种实验处理下要有相当的被试量(比如每一组被试是 20 人，就需要 960 人)，实验的操作简直不敢想象。这就是实际研究中，真正使用多因素完全随机实验设计的研究者很少的重要原因之一。查阅近几年国内发表的研究报告，你就会发现，《心理学报》、《心理科学》等刊物上难得找到几篇完全随机实验设计的研究，大部分研究采用多因素重复实验设计，其次是混合实验设计。

二、多因素重复实验设计

多因素重复实验设计(multi-factor repeated-measure design)也属于前一章讨论过的组内设计,多个自变量结合而构成的所有实验处理均由一个被试组来完成。比如,有三个自变量,其水平数分别是 p、q、r,则其结合而形成的实验处理数等于三者乘积 $p\times q\times r$。很显然,这种实验设计需要的被试数较少,因此带进实验的被试间差异也较少。当实验中的自变量都适合做被试内变量,且实验任务较简单,每次实验不花费很多时间,就可以使用多因素重复实验设计。这种实验设计在实际研究中使用很多,可以很容易地从《心理学报》和《心理科学》等专业刊物上找到这种实验设计的例子。

如陈燕丽等①采用 4×4 的重复实验设计对阅读四字成语时最佳的注视位置进行了实验研究。其研究是这样进行的:研究者从《成语大辞典》中选择了 32 个 4 类成语,其中 A 型成语是前面两个字一样,后面两个字一样,如“轰轰烈烈”;B 型成语是前面两个字不一样,后面两个字一样,如“目光炯炯”;C 型成语是前面两个字一样,后面两个字不一样,如“津津有味”;D 型成语是第一和第三个字一样,第二和第四个字不一样,如“古色古香”。然后编造了 32 个假成语,共 64 个真假成语构成了她们的实验材料。在电脑屏幕上呈现这些真假成语,让被试对这些真假成语作出“是”或“否”的判断。在每次呈现刺激材料前都要在屏幕上呈现一个注视位置提示标志“+”300 ms,“+”出现的位置对应于成语的四个字中的某一个字,每次出现的位置是随机的,而且在每个字位置上出现的次数相等。在出现成语或假成语时,要求被试通过按键作出“是”或“否”的判断,记录其反应时间和正确性。实验结束后分析成语类型不同、材料呈现前被试注视点位置不同对其判断速度和正确率有无影响。因为每个被试都完成上述所有 64 个刺激的实验任务,属于典型的 4×4 重复实验设计,自变量为两个(成语类型和刺激呈现前被试注视点的位置)。研究中虽然只使用了 33 名大学生,但因为每个被试完成了所有实验处理的实验任务,保证了每种实验操作条件下一组数据的样本容量,提高了研究的准确性。

① 陈燕丽、史瑞萍、田宏杰:《阅读成语时最佳注视位置的实验研究》,《心理科学》,2004, 27(2): 278—280。

为了说明重复实验设计数据的分析过程，我们现在假定本节前面谈到的错觉研究实验采用的是重复实验设计：如表 2-3 所示，研究者从某大学文学院本科二年级一班级随机抽取了 5 名男生，每一被试均接受全部四种实验处理。那么如何分析如表 2-3 所示的实验数据呢？

表 2-3　箭头方向与箭头角度对缪勒-莱尔错觉量的影响

被试	箭头方向向外(A_1)		箭头方向向内(A_2)		$\sum$
	箭头张开 15°(B_1)	箭头张开 45°(B_2)	箭头张开 15°(B_1)	箭头张开 45°(B_2)	
1	6	4	8	7	25
2	5	3	7	6	21
3	7	5	9	7	28
4	6	4	8	6	24
5	7	5	9	8	29
$\sum$	31	21	41	34	127

这一数据分析的目的也是考察两个自变量及其交互作用对缪勒-莱尔错觉量的影响。根据统计学方法，拟采用多因素重复实验设计的方差分析来确定自变量的效应。这一方差分析的过程如下：

第一步：计算数据总变异量并对之进行分解

表 2-3 中数据变化的原因大致可以划分为五个方面：(1)自变量 A 的主效应；(2)自变量 B 的主效应；(3)A 和 B 的交互效应；(4)被试差异带来的影响；(5)其他随机变量的影响，即残差。就本例来说，变异量 SS_T、SS_A、SS_B 和 SS_{AB} 各项的计算过程与前述完全随机实验设计方差分析中各项计算的过程完全一样，因此得到：

$$SS_T = 52.55$$

$$SS_A = 26.45$$

$$SS_B = 14.45$$

$$SS_{AB} = 0.45$$

与完全随机实验设计不同的是，在多因素重复实验设计的实验中，每一被试均参加了多种实验条件下的实验，得到了两个以上的观测结果，这样就可以将全部实验数据按照被试的不同进行分组。在表 2-3 中，共有 5 名被试参加实验，因

此 20 个数据可以按被试差异分成 5 组(即表中的 5 行),计算这 5 组数据之间的变异性,就可以估算被试差异带来的影响,并将这一影响引起的数据变异量从残差中分离出来。类似于前述变异量的计算方法,被试间差异引起数据的变异量可计算如下:

$$SS_S = \sum(X-\bar{X})^2 = \sum X^2 - (\sum X)^2/N$$
$$= \sum_j \frac{(\sum_i X_{ij})^2}{4} - (\sum\sum X_{ij})^2/N$$
$$= (25^2/4 + 21^2/4 + 28^2/4 + 24^2/4 + 29^2/4) - 806.45 = 10.30$$

于是,残差引起的变异量为:

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} - SS_S$$
$$= 52.55 - 26.45 - 14.45 - 0.45 - 10.30 = 0.90$$

第二步:计算各项数据变异对应的自由度

总变异对应的自由度:$df_T = 20 - 1 = 19$。然后将自由度分解:

A 的主效应对应的自由度:$df_A = a - 1 = 1$ (a 是自变量 A 的水平数)

B 的主效应对应的自由度:$df_B = b - 1 = 1$ (b 是自变量 B 的水平数)

A 和 B 交互效应对应的自由度:$df_{AB} = (a-1)(b-1) = 1$

被试的差异效应对应的自由度:$df_S = n - 1 = 4$

残差引起数据变异的自由度等于总的自由度减去上述四项:$df_E = 19 - 7 = 12$

第三步:计算各有关变异源引起数据变异的方差,即均方 *MS*

$MS_A = SS_A/df_A = 26.45$

$MS_B = SS_B/df_B = 14.45$

$MS_{AB} = SS_{AB}/df_{AB} = 0.45$

$MS_S = SS_S/df_S = 2.575$

$MS_E = SS_E/df_E = 0.90/12 = 0.075$

第四步:计算各效应是否显著的检验统计量 *F* 比率

$F_A = MS_A/MS_E = 26.45/0.075 = 352.667$　分子与分母的自由度为(1, 12)

$F_B = MS_B/MS_E = 14.45/0.075 = 192.667$　分子与分母的自由度为(1, 12)

$F_{AB} = MS_{AB}/MS_E = 0.45/0.075 = 6.000$　分子与分母的自由度为(1, 12)

第五步:给出方差分析表和分析结论,如表2-4所示

查F表确定各效应F比率达到统计学上的显著性水平所需的临界值,得到:

$$F(1,\ 12)\mid_{\alpha=0.05}=6.55\quad F(1,\ 12)\mid_{\alpha=0.01}=11.75$$

将上述F比率与临界值比较,就可确定各效应的F比率是否达到显著性水平的要求。从比较的结果知道:F_A和F_B均大于$F(1,\ 12)\mid_{\alpha=0.01}$,但$F_{AB}$小于临界值$F(1,\ 12)\mid_{\alpha=0.05}$和$F(1,\ 12)\mid_{\alpha=0.01}$。将上述分析的结果汇总,如表2-4所示。

表2-4　箭头方向与箭头角度对错觉量影响的方差分析表

变异来源	平方和	自由度	均方	F	p
A的主效应	26.45	1	26.45	352.667	<0.01
B的主效应	14.45	1	14.45	192.667	<0.01
AB的交互效应	0.45	1	0.45	6.000	>0.05
被试间效应	10.30	4	2.575		
误　差	0.90	12	0.075		
合　计	52.55	19			

从方差分析表可以看出,自变量A和自变量B的主效应达到了显著性水平($p<0.01$),A和B的交互效应没有达到显著性水平($p>0.05$)。因此,也可以得到结论:箭头的方向和张开角度对缪勒-莱尔错觉量有显著性影响,且二者对错觉量的影响是相互独立的(***提醒:此由假想的数据得到的结论,非真实研究的结论***)。

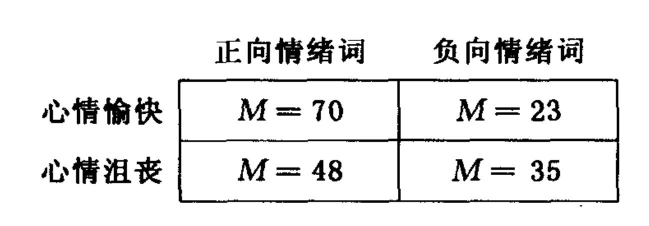
很明显,重复实验设计的方差分析中,可以将被试差异带来的数据变异从误差项中分离出来,使自变量的效应更容易显示出来,再加上这种实验设计非常节省被试,成为最常用的实验设计方法就不为怪了。但是,当实验的顺序效应比较明显、实验任务较大以致形成被试不可接受的负担时,就会在实验进程中引起新的混淆变量,造成研究内部效度的降低,则不宜采用重复实验设计。

三、混合实验设计

在多因素实验研究中,研究者经常会遇到这样的情况:在拟考察的几个自变量中,有一个或几个变量适合被试间设计,另一个或几个变量适合被试内设计;

或者,研究者更喜欢用被试内设计,以便使用较少的被试,但研究中有些自变量引出的不同实验条件可能存在较大的顺序效应,所以该自变量宜采用被试间设计。这时,就需要采用混合实验设计(mixed experimental design)方法来完成研究,该设计中同时包含被试间因子和被试内因子,用矩阵表示的话,则可以用被试间因子构成行,被试内因子构成列,每一行对应的一个或一组被试必须接受列的所有不同的实验处理。

混合实验设计"是一种将两种不同的研究策略结合在一起的因素型研究方法,如将被试间设计和被试内设计结合,或将一个实验因子与一个非实验因子结合"。[①] 在二因素实验中,如果一个自变量采用组间设计,另一个自变量采用组内设计,就构成了最简单的混合实验设计。如图 2-3 所示的实验设计就是一个最简单的二因素混合实验设计。[②]



	正向情绪词	负向情绪词
心情愉快	$M=70$	$M=23$
心情沮丧	$M=48$	$M=35$

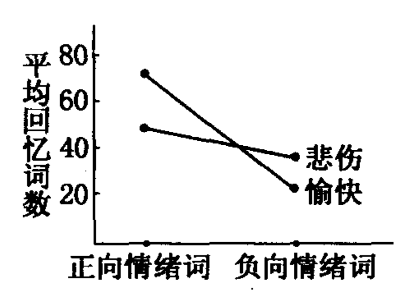


图 2-3　由一个被试间因子和一个被试内因子构成的二因素混合实验设计的假设结果

研究者在一个组引起愉快情绪,在另一个组引起沮丧情绪,从而创设了一个被试间因子(愉快/悲伤)。然后,检测每组被试回忆正向情绪词和负向情绪词的数量,这创设了一个被试内因子(正向/负向)。

图 2-3 显示的是一个考察情绪与记忆之间关系的混合实验设计。在这一课题领域中,最具代表性的研究结果表明:人们倾向于回忆那些与他们当前的情绪一致的信息。因此,心情愉快时,人们就容易回忆出快乐的事情,心情沮丧时人们就容易回忆出悲伤的事情。在一项如图 2-3 所示的研究中,蒂斯代尔和福加蒂(Teasdale & Fogarty, 1979)[③]通过两种方式操纵情绪,其中一组被试读一系

① [美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著,邓铸、姜子云、蒋小慧等译:《行为科学研究方法》,陕西师范大学出版社 2005 年版,第 315 页。

② 同上,第 211—212 页。

③ Teasdale, J. D. & Fogarty, S. J. (1979). Defferential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, 88: 248—257.

列越来越悲伤沉闷的陈述(例如“回首我的生命历程,我怀疑是否获得过任何真正有价值的东西”);另一组被试读一系列越来越轻松、高兴的陈述(例如“生活如此充实、有趣,活着真是太棒了”)。这样,研究者就创设了一个被试间实验因子,包括情绪高兴组和情绪沮丧组,这两个组分别对应于矩阵中的两行。然后向所有被试呈现一个词单,其中包括一些积极的、令人愉快的词汇,也包括一些消极的、令人沮丧的词汇。最后,研究者分别记下每位被试能回忆的正向情绪词数和负向情绪词数。这里,研究者又创设出一个被试内因子,包括正向情绪词和负向情绪词,它对应于矩阵中的两列。

当然,混合实验设计的含义不仅仅是指被试间设计与被试内设计的混合,也包括实验的与准实验的混合、实验的与非实验的混合、准实验的与非实验的混合等。

行为科学研究中,采用由一个实验因子和一个准实验因子构成的因素型实验设计是很常见的。在这样的混合实验设计中,有一个因子是真正的自变量,它包括一系列可被操纵的实验条件;另一个因子是准自变量,主要有两类:①

1. 现成的被试特征,如年龄或性别等。研究者想考察实验处理条件对男性和女性是否会产生同样的效应,或者想知道实验处理的效应是否会随着年龄的变化而变化。现成被试特征将被试自然分成两组或多组,因此它是一个准实验因子。

2. 第二个因子是时间。如研究者关注不同实验处理的效应会持续多长时间。比如,两种不同治疗技术在治疗结束后会立即产生相同的作用,但经过一段时间后,其中一种治疗能继续产生效应,而另一种治疗的效应随着时间推移而逐渐消失。在这类研究中,时间因素是不受研究者控制或操纵的,因此它也是一个准实验因子。

什拉格(Shrauger, J. S., 1972)曾对人的行为是否会受观众影响的问题进行研究。他安排被试完成一项概念形成的实验任务,其中有一半被试是独自工作的(无观众),另一半被试则是在有观众的情境中工作的,而且观众都表现出对被试操作成绩的极大兴趣。实验中,有无观众的安排是由研究者掌控的,所以这是一个实验因子。第二个因子是自尊水平,根据研究者事先进行的测量,被试被

---

① [美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著,邓铸、姜子云、蒋小慧等译:《行为科学研究方法》,陕西师范大学出版社2005年版,第213页。

划分为高自尊组和低自尊组。自尊水平的高低是被试本来已有的机体特征,因此它是一个准实验因子。实验结果如图 2-4 所示,两个因子间有明显的交互效应。具体地说,观众在场对低自尊的被试有明显作用,但对于高自尊的被试来说几乎没有什么影响。①

	无观众	有观众
高自尊	$M = 2.1$	$M = 2.2$
低自尊	$M = 6.1$	$M = 9.5$

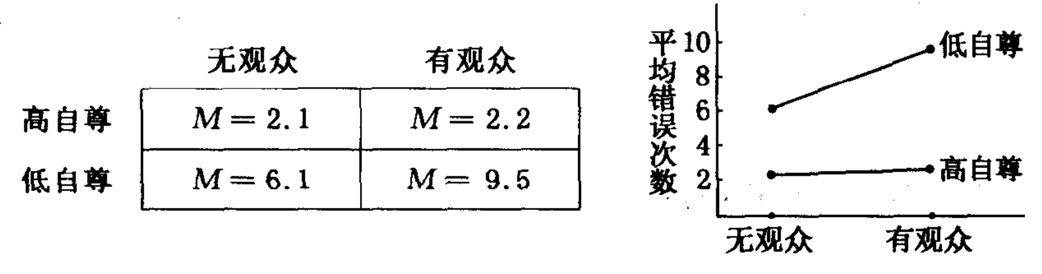


图 2-4　由一个实验因子和一个准实验因子结合构成的二因素混合实验设计的假设结果

为了创设一个实验因子(有无观众),研究者可以控制被试是或不是在有观众条件下操作任务;为了创设一个准实验因子,研究中观测两个不等的被试组(自尊水平的高低)。实验中的因变量是每个被试操作中的错误次数。

我们在前文已经指出,国内心理学家的研究中,除重复实验设计之外,就数混合实验设计的使用频率高了。有兴趣的读者可以在《心理学报》、《心理科学》等刊物上查阅一些混合实验设计的研究实例。

第三节　随机区组实验设计与拉丁方实验设计

已经学习过完全随机实验设计,也就是将被试随机分成相等的多组,各组被试各自完成一种实验处理,然后比较各组被试因变量观测值的差异。研究者可以使用 t 检验(两个样本组)、Z 检验(两个大样本组)或方差分析(两个以上的样本组)对平均数差异的显著性水平进行检验,方差分析最能体现差异检验的逻辑。方差分析是将所有观测值的总体变异分解为自变量的主效应、自变量的交互效应(当自变量不止一个时)、误差效应(包括组内被试的差异和其他因素导致的变异,可以统称为残差),然后将自变量效应、交互效应的方差与误差效应方差比较求得 F 比率。F 比率越大,表明自变量或自变量的交互效应越明显。但在计算 F 比率时,误差效应方差是作为分母出现的,也就是说误差项越小,F 比率

① 转引自:[美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著,邓铸、姜子云、蒋小慧等译:《行为科学研究方法》,陕西师范大学出版社 2005 年版,第 213 页。

就会越大,对自变量效应的检验就越敏感。

这样的分析启发我们,方差分析中的残差越大越是不能显示出自变量的影响效应。如果在实验设计中能减少未知因素带来的误差项即残差,方差分析就能更灵敏地将自变量的效应显示出来。由此,我们介绍随机区组实验设计(randomized block experimental design)和拉丁方实验设计(Latin square experimental design)及其数据的分析过程。

一、单因素随机区组实验设计

在行为科学研究中,接受处理的实验单位一般都是一个人、一只老鼠或其他动物。几乎可以肯定地说,对于我们想要测量的任何变量来说,个体之间总会存在差异性,而未经选择的一组被试之间的差异就更大。比如,他们在反应时间、问题解决能力、学习能力、记忆能力等方面都会有所不同。因此,实验中的个体差异必然会带来因变量测量数据的变异,当这些变异无法从残差中分离出来时,方差分析的敏感度就大大降低了,自变量的效应也因此较难显现出来。这里讨论的随机区组实验设计,为将个体差异带来的数据变异从残差中分离出来提供了有效方法。研究中,可以利用机体特征将被试划分成若干区组,并将他们因变量的观测值分开记录,这样就可以计算不同区组之间因变量的变异量,从而将由于个体差异带来的数据变异从残差项中分离出来,达到降低残差项的目的。虽然更多时候,研究者是将机体变量作为区组变量,但区组变量也可以是其他刺激变量,如不同测试时间、不同测试地点、不同测试实施者等。

现以单因素随机区组设计为例来说明区组设计的模式及数据分析方法。假如,某研究者想考察不同箭头角度对缪勒-莱尔错觉的影响,从大学生中抽取了一些被试参加实验,但这些学生分别来自数学系、化学系、中文系和考古系。考虑到这些学生接受的不同专业训练可能会造成其缪勒-莱尔错觉量有较大的个体差异,于是决定采用单因素随机区组实验设计方法来完成这一研究,将被试接受的不同专业训练作为一个区组变量。现有 36 名被试参加缪勒-莱尔错觉实验,自变量为箭头角度,有三个水平,分别为 30°、45°和 60°。36 名被试中来自数学系、化学系、中文系、考古系的各 9 人,每个专业的学生都被随机均分到自变量的三个水平上,以此实现对三种实验条件下被试专业训练因素的平衡。

这一设计不仅考虑了自变量的影响,而且也考虑了被试本身某一特征的影

响。将被试间的某种差异因素作为一个区组变量，至少可以部分地把被试间差异引起的因变量的数据变异从残差中分离出来。需要说明的是，区组变量不是自变量，而是额外变量，因为它不是研究者拟考察的，而是研究者要平衡的。一般作为区组变量的额外变量与自变量之间不存在交互作用，如果存在交互作用，它就不适宜作为区组变量，而应作为自变量来加以研究，或作为额外变量要加以控制。

假设上述实验设计得到的观测数据如表 2-5 所示。

表 2-5　箭头角度与缪勒-莱尔错觉量的关系

区　组	30°	45°	60°	$\sum$
数学系	6 5 5	5 4 5	3 4 3	40
考古系	7 7 6	6 5 4	4 5 3	47
化学系	7 9 8	7 6 7	4 6 5	59
中文系	8 7 8	7 6 8	4 5 7	60
$\sum$	83	70	53	206

对于这一实验设计模型，其数据如何处理呢？很显然，如果不去关注不同专业学生(区组)之间的差异，这一实验就是一个单因素完全随机实验设计，对其进行单因素方差分析来考察自变量(缪勒-莱尔错觉实验中角度的张开度数)的影响效应，这时不同专业的差异带来的数据变异就与其他随机变量带来的变异混淆在一起，作为残差项。如果把一个专业的学生作为一个区组，就可以计算一个区组变量对测量错觉量所带来的变异平方和，将其从残差项平方和中分离出来，使残差项方差降低，自变量的效应也因此就更容易显示出来了。

完全随机区组实验设计的数据分析与完全随机实验设计的数据分析相比，就是要多计算一个变异平方和——区组变量引起因变量变异的平方和。当然，区组变异的自由度也可以计算出来，它等于区组数减 1。在将区组变量引起的

变异从残差中剔除时,也要将区组变量的自由度从残差项自由度中减除。实际上,这里可以检验区组变量的效应是否显著,方法类似于自变量效应的检验。

区组实验设计要求区组变量与自变量不存在交互效应,即二者的交互效应不应达到显著性水平。但是,我们仍然主张对区组变量与自变量的交互效应引起的变异量及自由度进行计算和检验,这样做至少有两点好处:首先,计算出这一交互效应引起的因变量的变异平方和,就可以将其从残差项中减除,起到减少残差项的作用;其次,对这一交互效应是否显著进行检验,本身也是对区组实验设计适宜度的检验。如果检验发现这一交互效应显著,就说明采用区组设计是不适当的,需改用其他设计重新研究。本例方差分析中的误差平方和等于总变异平方和减去自变量变异平方和、区组变量变异平方和、自变量与区组变量交互效应变异平方和三部分。①由此可见,在进行方差分析时,就暂且把区组变量作为自变量看待了。

区组实验设计一般是在被试选取或分组过程中,有些被试变量不易控制时使用的。而且需要注意的是,在进行区组实验设计的时候,每个区组中的被试数必须是实验处理数的整数倍,这样才能保证将每个区组中的被试均分到各个实验处理当中去。

有些时候,每个区组中的被试只分配到每种实验处理水平上一个被试,数据分析的过程也和上述讨论的过程一样。实际上,这种情况与我们以前讨论到的配对组设计相同。

最后,需要对单因素重复实验设计、单因素配对实验设计、单因素完全随机区组实验设计作一比较。首先,单因素重复实验设计与单因素配对组实验设计在本质上是一致的,因为配对组之间可被看作是相互映射或相互替代的,所以配对组设计的数据分析方法与重复设计的数据分析方法相同,只不过两种设计使用的被试数不同,后者使用的被试数是前者的 k 倍,k 就是实验处理数。其次,单因素配对实验设计可看作是单因素完全随机区组实验设计的一个特例,即当每一区组的被试数正好等于实验处理数时,每一区组中的被试分到每一实验条件中的只有一名被试,这时的区组设计也就是匹配设计,其在本质上也就同于重复实验设计。最后,单因素重复实验设计与单因素区组实验设计是两种不同的实验设计,因为重复实验设计中只有一组被试,而区组实验设计中是有多组被试

① 对于区组变量与自变量的交互效应感兴趣的读者可查阅:[美]Allen L. Edwards 著,毛正中等译:《心理研究中的实验设计(第五版)》,四川教育出版社 1996 年版,第 381—439 页。

的,且多数情况下,每一区组被试分配到每一实验条件中的被试也不只有一名,况且区组变量也不都是被试变量,它还可以是刺激变量等。

二、多因素随机区组实验设计

当研究的自变量不止一个,同时考虑一个区组变量的时候,就需要采用多因素随机区组实验设计。下面以假想的例子来说明。

如为试验两种新教材与两种教法的匹配关系,研究者同时考虑到学生现有成绩水平的影响而将其作为一个区组变量,选择优、中、差各 12 个班进行试验。经过一学年的教学过程后进行学生学习成绩的比较,即年终各班平均的考试成绩如表 2-6 所示。

表 2-6 教材教法实验研究结果

区组	教材 1(A_1)		教材 2(A_2)		$\sum$
	教法 1(B_1)	教法 2(B_2)	教法 1(B_1)	教法 2(B_2)	
优等班(C_1)	50 40 50	60 50 70	50 40 60	80 70 90	710
中等班(C_2)	40 40 30	50 50 40	40 50 50	70 60 60	580
差等班(C_3)	20 30 30	50 40 40	30 40 30	50 40 50	450
$\sum$	330	450	390	570	1 740

现在以表 2-6 中数据为例说明多因素随机区组实验设计的方差分析过程:

第一步:计算数据总变异量并对之进行分解

表 2-6 中数据变化的原因大致可以划分为五个方面:(1)自变量 A 的主效应;(2)自变量 B 的主效应;(3)A 和 B 的交互效应;(4)区组变量 C 的主效应;(5)其他随机变量的影响(其中也包括一定量的区组变量与自变量的交互效应,此处不再对其进行计算),即残差。

根据前述的方差分析中变异量的计算方法，得到：

$SS_T = 7\,900$

$SS_A = [(330+450)^2/18+(390+570)^2/18]-84\,100 = 900$

$SS_B = [(330+390)^2/18+(450+570)^2/18]-84\,100 = 2\,500$

$SS_{AB} = (330^2/9+450^2/9+390^2/9+570^2/9)-SS_A-SS_B-84\,100 = 100$

$SS_C = (710^2/12+580^2/12+450^2/12)-84\,100 = 2\,816.67$ （区组变量引起的变异平方和）

$SS_E = SS_T-SS_A-SS_B-SS_{AB}-SS_C = 7\,900-900-2\,500-100-2\,817.67 = 1\,583.33$

第二步：计算各种效应引起数据变异的自由度（即数据发生变异的次数）

总的数据变异对应的自由度：$df_T = 36-1 = 35$。然后将自由度分解：

A 的主效应对应的自由度：$df_A = a-1 = 1$ （a 是自变量 A 的水平数）

B 的主效应对应的自由度：$df_B = b-1 = 1$ （b 是自变量 B 的水平数）

A 和 B 交互效应对应的自由度：$df_{AB} = (a-1)(b-1) = 1$

区组变量 C 的主效应对应的自由度：$df_C = c-1 = 2$ （c 是区组数）

残差对应的自由度等于总的自由度减去上述四项：$df_E = 35-5 = 30$

第三步：计算各变异源引起数据变异的方差，即均方 *MS*

$MS_A = SS_A/df_A = 900/1 = 900$

$MS_B = SS_B/df_B = 2\,500/1 = 2\,500$

$MS_{AB} = SS_{AB}/df_{AB} = 100/1 = 100$

$MS_C = SS_C/df_C = 2\,816.67/2 = 1\,408.34$

$MS_E = SS_E/df_E = 1\,583.33/30 = 52.78$

第四步：计算各效应是否显著的检验统计量 *F* 比率

$F_A = MS_A/MS_E = 900/52.78 = 17.052$　分子与分母的自由度为(1, 30)

$F_B = MS_B/MS_E = 2\,500/52.78 = 47.366$　分子与分母的自由度为(1, 30)

$F_{AB} = MS_{AB}/MS_E = 100/52.78 = 1.895$　分子与分母的自由度为(1, 30)

$F_C = MS_C/MS_E = 1\,408.34/52.78 = 26.683$　分子与分母的自由度为(2, 30)

第五步：给出方差分析表和分析结论，如表 2-7 所示

查 F 表确定各效应 F 比率达到统计学上的显著性水平所需的临界值，得到：

$$F(1,\ 30)\mid_{\alpha=0.05} = 5.57 \quad F(1,\ 30)\mid_{\alpha=0.01} = 9.18$$

$$F(2,\ 30)\mid_{\alpha=0.05} = 4.18 \quad F(2,\ 30)\mid_{\alpha=0.01} = 6.35$$

将上述 F 比率与临界值比较，就可以确定各效应的 F 比率是否达到显著性水平。从比较的结果知道：F_A 和 F_B 均大于 $F(1, 30)|_{\alpha=0.01}$，但 F_{AB} 小于临界值 $F(1, 30)|_{\alpha=0.05}$ 和 $F(1, 30)|_{\alpha=0.01}$；F_C 大于 $F(2, 30)|_{\alpha=0.01}$。将上述分析的结果汇总，如表 2-7 所示。

表 2-7　教材教法实验研究结果的方差分析表

变异来源	平方和	自由度	均方	F	p
A 的主效应	900	1	900	17.052	<0.01
B 的主效应	2 500	1	2 500	47.366	<0.01
AB 的交互效应	100	1	100	1.895	>0.05
区组变量主效应	2 816.67	2	1 408.34	26.683	<0.01
误　差	1 583.33	30	52.78		
合　计	7 900	35			

分析的结果显示，两个自变量的主效应都达到了显著性水平($p < 0.01$)，但二者的交互效应未达到显著性水平($p > 0.05$)。区组变量的主效应也达到了非常显著性的水平，表明该实验设计采用随机区组实验设计是非常必要的，它将学生现有学习成绩的差异影响的大部分都从残差项中分离出来，自变量的效应更容易显示出来。

三、拉丁方实验设计

区组实验设计是在考察自变量影响效应的实验中，考虑到一个额外变量的影响，将这个额外变量作为区组变量，对其在各种实验处理条件下产生的影响进行平衡，同时将该区组变量引起的变异从残差中分离出来。如果将区组实验设计进一步扩展，即考虑两个额外变量的影响，而欲对这两个额外变量的影响进行平衡并将其引起的变异从残差项中分离出来时，就需要采用拉丁方实验设计。与随机区组实验设计相比，拉丁方是其扩展形式。

比如，限于实验室条件，研究者开展某一实验研究时每天只能为四名被试进行测试，实验处理也有四个水平：A_1、A_2、A_3、A_4。如果认为不在同一天进行测试，可能会引起测试结果的变化，这种影响又是比较重要的。于是可以将测试时间作为区组变量，即把同一天接受测试的被试看作是一个区组。这样就可以形成一个区组实验设计，如表 2-8 所示。

表 2-8 四种实验处理的随机区组实验设计

区　组	A_1	A_2	A_3	A_4
第一天				
第二天				
第三天				
第四天				

假如,在每天的实验中,一次只能测试一人,每天参加实验的四名被试只能分别在下午 2～3 点、3～4 点、4～5 点和 5～6 点的四个时段接受测试,而测试时段不同也可能会造成结果变化。这样一来,每一种实验处理条件安排的时段就也要取得平衡才行,你不能每天都在 2 点钟安排所有被试接受 A_1 处理条件,或 3 点钟接受 A_1 处理条件。于是,研究中采用测试日期和测试时段两方面因素的平衡方法安排实验,构成了一个单因素的拉丁方实验设计,设计模式如图 2-9 所示。在这一设计中,测试日期和测试时段在各种实验处理之间取得了很好的平衡。

表 2-9 四种实验处理的拉丁方实验设计

日　期	2～3 点	3～4 点	4～5 点	5～6 点
第一天	A_1	A_2	A_3	A_4
第二天	A_2	A_3	A_4	A_1
第三天	A_3	A_4	A_1	A_2
第四天	A_4	A_1	A_2	A_3

从上述例子可以看出,拉丁方(latin square)是一个含 P 行 P 列,把 P 个实验处理分配给 $P \times P$ 方格的管理方案,它便于在复杂研究程序中有条理地管理各个工作单元,并平衡两种额外变量的影响。在工农业生产试验和心理与教育研究中,拉丁方都得到普遍应用。在这种实验设计中,首先根据自变量处理的水平数确定两个额外变量的水平数,然后利用两个额外变量的各个水平结合在一起构造一个方格,最后再将自变量的不同处理平衡地安排在这个方格中,就构成了一个研究方案,其结果要保证自变量的每一个水平在拉丁方格的每一行和每一列都出现且只出现一次。很明显,在这种设计中,自变量的水平数或水平结合数、额外变量的水平数必须相等。

拉丁方设计常被用于平衡实验安排的时空顺序,也可被用于平衡机体变量

的影响。我们再以下面这个例子对拉丁方做进一步说明。

问题模式：

为了研究简单反应时间与光刺激的颜色和强度的关系，研究者同时考虑到被试的气质类型及年龄因素可能对反应时间具有明显影响，为了将这两个因素的影响从变异的残差项中分离出去，研究者采用了拉丁方实验设计。

拉丁方格的组成：

拉丁方格是由实验中明显存在的两个额外变量即被试的气质类型和被试年龄档组成，其中年龄分为四档：10～13 岁、15～18 岁、20～23 岁、25～28 岁。从四个年龄档的青少年中筛选出四种典型气质类型者各 2 人，这样就有共计 32 名被试参加这一实验。根据气质类型和年龄档组成拉丁方格，拉丁方格中的每一个格子中可以有年龄档相同、气质类型相同的 2 名被试，如表 2-10 所示。

表 2-10　4×4 拉丁方格

		被试气质类型			
		多血质	*胆汁质*	*黏液质*	*抑郁质*
被试年龄档	*10～13*				
	15～18				
	20～23				
	25～28				

实验处理的组成：

实验中自变量有两个，即光的颜色和强度。自变量的颜色取两个水平，红光和绿光，分别用 A_1 和 A_2 表示；光的强度也取两个水平，相对强度为 1 和 1/4，分别用 B_1 和 B_2 表示。于是两个自变量结合而成的实验处理分别为：

A_1B_1——红光＋1（即光的颜色为红光、光的相对强度为 1）

A_1B_2——红光＋1/4

A_2B_1——绿光＋1

A_2B_2——绿光＋1/4

实验处理的编排：

按照拉丁方实验设计的基本原则，将四种实验处理安排在拉丁方格中，某种实验处理被分配到拉丁方格中的某一方格，该方格中对应的两个被试就要完成这一种实验处理。

首先，我们给出一个基本的拉丁方设计形式，如表 2-11 所示。

表 2-11　标准的 4×4 拉丁方实验方案

被试年龄档次	多血质	胆汁质	黏液质	抑郁质
10～13	A_1B_1	A_1B_2	A_2B_1	A_2B_2
15～18	A_1B_2	A_2B_1	A_2B_2	A_1B_1
20～23	A_2B_1	A_2B_2	A_1B_1	A_1B_2
25～28	A_2B_2	A_1B_1	A_1B_2	A_2B_1

（列：被试气质类型）

表 2-11 所示的实验设计方案就是一个标准的或基本的 4 × 4 拉丁方的实验设计。有了这样的设计方案之后，实验程序的编排就非常清晰了。按照这一设计进行实验，不仅能将两个额外变量的效应从残差项中分离出来，而且有利于增进复杂实验过程的条理性。有了表 2-11 所示的实验方案，每个被试需要完成什么样的实验就很清晰了，比如 15～18 岁组两个胆汁质的学生只需完成 A_2B_1 实验处理，即"绿光＋1"实验处理、25～28 岁组两个黏液质的学生只需完成 A_1B_2 实验处理，即"红光＋1/4"实验处理。

有了表 2-11 所示的标准拉丁方实验设计方案之后，还可以将该方案进行随机化处理，即可以对其中的实验安排做随机的两行互换或两列互换，得到各种不同的拉丁方实验方案。比如，将表 2-11 中第 1 列和第 4 列对换就可以得到表 2-12所示的拉丁方实验方案。

表 2-12　在标准 4×4 拉丁方实验方案基础上变换得到的实验方案

被试年龄档	多血质	胆汁质	黏液质	抑郁质
10～13	A_2B_2	A_1B_2	A_2B_1	A_1B_1
15～18	A_1B_1	A_2B_1	A_2B_2	A_1B_2
20～23	A_1B_2	A_2B_2	A_1B_1	A_2B_1
25～28	A_2B_1	A_1B_1	A_1B_2	A_2B_2

（列：被试气质类型）

再将表 2-12 中的第 2 行和第 3 行对换就可以得到表 2-13 所示的拉丁方实验方案。

表 2-13　在表 2-12 基础上变换得到的拉丁方实验方案

被试年龄档 \ 被试气质类型	多血质	胆汁质	黏液质	抑郁质
10～13	A_2B_2	A_1B_2	A_2B_1	A_1B_1
15～18	A_1B_2	A_2B_2	A_1B_1	A_2B_1
20～23	A_1B_1	A_2B_1	A_2B_2	A_1B_2
25～28	A_2B_1	A_1B_1	A_1B_2	A_2B_2

进行拉丁方实验设计中,其选取用来构成拉丁方格的额外变量不能与研究的自变量之间存在交互效应,两个额外变量之间也不能存在交互效应。其数据的方差分析方法与随机区组实验设计相似,可以对数据的变异及其自由度进行分解,计算过程是:首先计算总变异,然后计算自变量及其交互效应引起的变异、两个额外变量主效应引起的变异,再计算误差项变异,即可得到各种变异方差及其与误差方差的比率 F。

拉丁方实验设计既有优点也有缺点。其优点是,在许多研究情境中,这种设计比完全随机实验设计和随机区组实验设计更加有效,它可以使研究者平衡并分离出两个额外变量的影响,因而减小实验误差,可获得对实验处理效应更精确的估价。另外,通过对方格单元内误差与残差的 F 检验,可以检验额外变量与自变量是否有交互作用,以检验采用拉丁方设计是否合适。拉丁方设计的缺点是,它的关于自变量与额外变量不存在交互作用的假设在很多情况下都难以保证,尤其当实验中含有多个自变量的时候。因此,拉丁方实验设计在多因素实验中不常用。另外,拉丁方实验设计要求每个额外变量的水平数与实验处理数必须相等,这也在一定程度上限制了拉丁方实验设计的使用。①

阅读材料 2-1

主效应与交互效应的关联性

在析因实验(多因素实验)中,数据收集、数据分析的主要目标是考察自变量

① 舒华:《心理与教育研究中的多因素实验设计》,北京师范大学出版社 1994 年版,第 58 页。

的主效应和交互效应是否显著。一个自变量的主效应显著，意味着该自变量的各个水平在其他自变量的所有水平上的平均数存在差异；否则，就不存在显著性差异。比如，在自变量A和自变量B构成的2×2析因设计中，如果A的主效应显著，那就意味着A_1在B_1和B_2水平下的平均数与A_2在B_1和B_2水平下的平均数存在显著性差异。变量间的交互效应则是指一个因子的效应依赖另一个因子的不同水平。

在析因设计中，方差分析直接给出自变量的主效应和交互效应是否显著的结果，多数研究者也依此判定自变量的作用是否明显、这些自变量的作用是否相互依赖。事实上，自变量的主效应与交互效应的评估并非这么简单，它们存在关联性，需要具体情况具体分析。我们就以两个自变量的主效应和交互效应来分析。当交互效应不显著的时候，两个自变量相互独立，我们可以直接从其主效应是否显著来评估自变量对因变量的作用大小；当两个自变量间的交互效应显著时，就不能简单地从主效应是否显著直接得出结论了。我们现在以交互效应显著为前提，来区分自变量A的主效应是否显著的三种情况：

第一，交互效应显著，A的主效应也显著，而且主效应方向与简单效应方向一致，如图2-5中的b图就属于这类情况。这种情况下，在自变量B的两个水平上，自变量A从A_1到A_2的变化引起的因变量的变化趋势一致，只是变化幅度不一致。

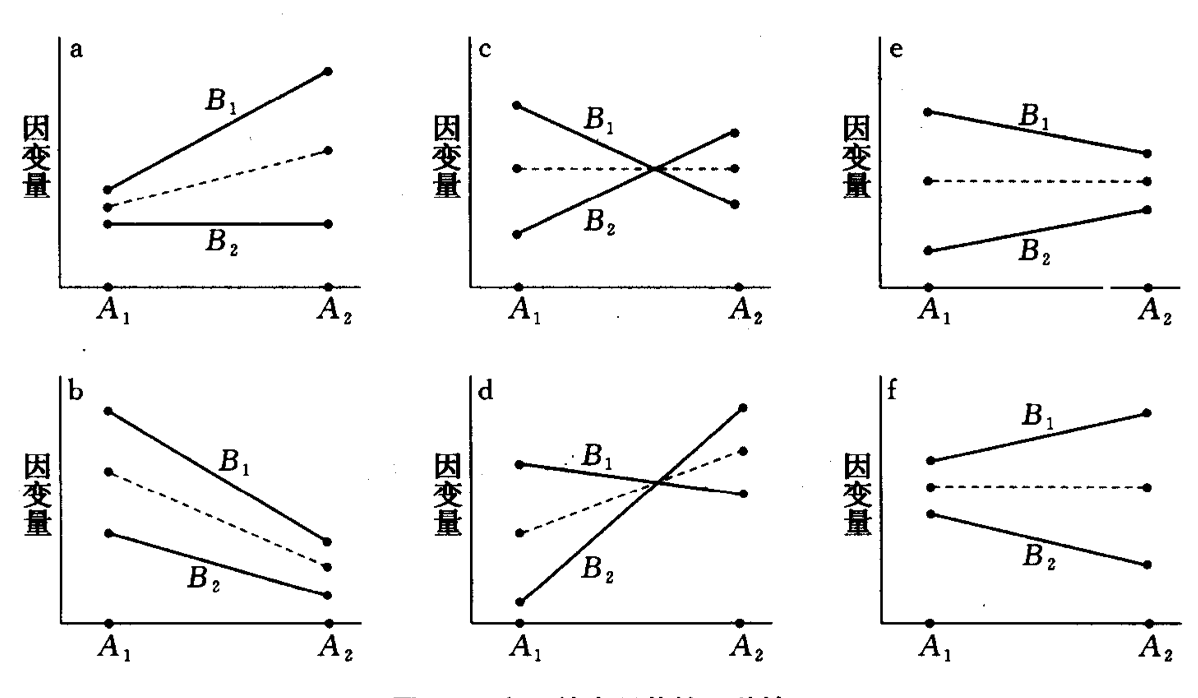


图2-5 交互效应显著的几种情况

这里的交互效应掩盖了自变量A在自变量B不同水平上的效应量的差异。很明显，在B_1上平上，A的效应量大于其在B_2水平上的效应量。

第二，交互效应显著，A的主效应也显著，这时A的效应方向可能会被交互效应歪曲。比如图2-5中的a图、d图都属于这类情况。在a图中，A的变化在B_1的水平上引起了因变量的显著变化，但在B_2水平上却未引起因变量的变化，这就是说A的变化不是在任何情况下都会引起因变量的变化，它依赖自变量B的水平；在d图中，虽然A的变化在B的两个水平上都引起了因变量的明显变化，但是变化的方向正好相反，从其主效应看，A的水平提高可以促进因变量分数的提高，但实际情况是，当A在B_1水平上提高时，反而会导致因变量分数的下降。所以在这种情况下，显著的交互效应掩盖或歪曲了自变量A的作用机制：它在B的不同水平上效应量是不同的。

第三，交互效应显著，A的主效应却不显著，实际上是交互效应掩盖了A的效应，如图2-5中的c、e、f图都属于这种情况。我们从这些图示中可以明显看到A的效应，但方差分析结果却会显示A的主效应不显著，这是因为A在B的两个水平上的效应方向相反，计算A的主效应时A_1和A_2的差异量被掩盖在了平均过程中。

那么，如何依据自变量主效应及其与其他自变量的交互效应来进行结果分析呢？这一点很简单：当方差分析结果显示A的主效应及A与其他自变量的交互效应都不显著时，则说明A的效应真的不明显；当方差分析的结果显示A的主效应不显著但A与其他自变量的交互效应显著时，则说明A其实对因变量有明显作用，即A的效应其实是存在的，只不过其效应的大小和方向依赖其他自变量的不同水平。

上述分析提醒我们，在说明方差分析结果时你要特别注意，如果因子间的交互效应达到了显著性水平，那么自变量的效应有可能会被歪曲或掩盖，也就是说，不能简单地依据其主效应是否显著来判断它是否对因变量有影响，而是要进行简单效应检验，分别考察其在其他自变量不同水平上的变化情况。否则，可能会得到错误结论。应该记住，一个因子的主效应是对其在另外一个因子所有不同水平下观测分数的平均而得到的，而这种平均的结果可能很难准确地反映每种具体实验处理的效应。

总之，交互效应可能会掩盖或歪曲两个因子中任何一个因子的主效应。因此，只要是交互效应达到了统计学上的显著性水平，你在就主效应问题作出结论

前都要仔细考察具体的数据变化。

——资料来源:《行为科学研究方法》(邓铸等,2005)①

阅读材料 2-2

何种实验设计,如何分析数据?

蔡厚德教授(2005)在一项研究中,欲考察汉字知觉过程中的脑半球单侧优势效应及相关特征,探测了汉字呈现视野(包括左与右;偏离中心的不同视角)、左手反应或右手反应对汉字知觉以及相应任务操作的影响。下面是其关于实验研究方法的表述。那么,该研究采用了何种实验设计?其数据应该如何分析呢?

1. 被试

大学生 120 名(男女各 60 名,年龄 20～22 岁),经利手十项标准检查均为右利手,且视力正常。他们被分层随机分为 6 组,每组 20 名(男女各 10 名)。根据比较刺激呈现的偏心视角和反应用手,6 组被试分别对应于 3.5°—左手、3.5°—右手、5°—左手、5°—右手、6.5°—左手、6.5°—右手的六种实验条件。

2. 材料

刺激由 8 个汉字大写数字(贰、叁、肆、伍、陆、柒、捌、玖)组成,笔画数6～9($M=8.75$)。叁和捌作为标准刺激在中间视野呈现,同时有一比较刺激在左视野或右视野呈现。根据数字奇—偶概念的同/异,6 组刺激对(叁—伍、叁—柒、叁—玖、捌—贰、捌—肆、捌—陆)为"同"判断,其余 6 组刺激对(叁—贰、叁—肆、叁—陆、捌—伍、捌—柒、捌—玖)为"异"判断。每一被试共进行 2 × 12 组刺激对的测试。每一汉字刺激上下占 3.5°视角,左右占 3°视角。

3. 仪器与测试

刺激呈现以及反应时间和错误百分数的记录由 Intel-586 计算机控制。测试时要求被试将下颚置于屏幕前的观测架上,两眼距屏幕 35 cm。每次测试时由被试用拇指按压空格键启动计算机在屏幕中心呈现一个"十"字,被试两眼需注意凝视出现的"十"字。1 秒钟后"十"字消失,同时在中间视野和左视野或中

① [美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著,邓铸、姜子云、蒋小慧等译:《行为科学研究方法》,陕西师范大学出版社 2005 年版,第 205—210 页。

间视野和右视野出现一对汉字数字刺激。被试须根据字对的奇—偶概念同/异，尽可能快地用左或右手的中指按压“同”键，或食指按压“异”键。数字对的呈现顺序是随机的，比较刺激在左右视野的呈现是半随机的，同/异判断结果按*ABBA*和*BAAB*的方式排序。数字对的呈现时间为50 ms。

——资料来源：《刺激的知觉辨认难度与大脑两半球间的分布式加工》(蔡厚德，2005)①

建议阅读文献

1. 杨治良主编：《基础实验心理学》，甘肃人民出版社1988年版，第20—52页。
2. 舒华编著：《心理与教育研究中的多因素实验设计》，北京师范大学出版社1994年版。
3. 金志成、何艳茹编著：《心理实验设计及其数据处理》，广东高等教育出版社2002年版，第115—159页。

复习思考题

1. 如何理解完全随机实验设计、重复实验设计、混合实验设计、区组实验设计、拉丁方实验设计、方差分析、变异源、残差、析因实验设计、交互效应？
2. 多因素实验设计的类型主要有哪些？
3. 如何进行单因素完全随机实验设计的数据分析？
4. 如何进行多因素完全随机实验设计的数据分析？
5. 如何进行重复实验设计的数据分析？
6. 如何进行区组实验设计？
7. 如何进行拉丁方实验设计？
8. 结合自己的兴趣、经验并查阅相关文献，提出一个研究课题，在研读与该课题有关的一些研究论文后，制订一套研究方案(采用多因素实验设计方法)。

① 蔡厚德：《刺激的知觉辨认难度与大脑两半球间的分布式加工》，《心理学报》，2005，37(1)：14—18。

第三章
生态学方法与准实验设计

本章内容提要

心理学研究中越来越强调人的心理的文化属性，重视研究方法的多元化，即主张采用多种方法和手段将人放回到较为真实的生活情境中进行研究，这是心理学研究中的生态化运动。在心理学的生态性研究范畴中，最典型的研究设计是准实验设计。准实验设计主要包括单组准实验设计和多组准实验设计两大类，本章介绍单组准实验设计中的“时间序列设计”、“相等时间取样设计”，多组准实验设计中的“不等组前测后测设计”、“不等组前测后测时间序列设计”，主要介绍这几种研究设计的基本模式、资料分析方法、优缺点等。

由于人文学科过分追随自然科学的研究取向，采用还原手法，在某种程度上忽视了人的整体性研究，忽视了社会文化及生活境遇对人的精神的影响，人文主义和后现代哲学思潮对此进行了严厉批评，导致了人文科学研究中两种文化的尖锐对立，即科学主义与人文主义的对立①。这种对立也明显地存在于心理学的阵营中，并导致了一种折中的趋势。近年来，在包括心理学在内的人文学科发生着一场生态化运动(ecological movement)。心理学中的生态化运动顺应了心理学回归生活的潮流，其方法论也会逐渐成为国内外心理学研究中的重要方法论。本章，专门讨论心理学中的生态化运动和准实验设计方法。

第一节　心理学中的生态化运动

心理学中的生态化运动是由于实验心理学的极端自然科学化追求引起的，

① Kimble, G. A. (1967). Psychology's two culture. *American Psychologist*, 39(8).

这种极端自然科学化在信息加工心理学中表现最为突出。早期信息加工心理学完全采用人机类比的思想方法，将人脑比作电脑，将人的心理活动机制还原成神经纤维上的电信号及其传递，丝毫不顾忌人的意识活动的内容效应和情境效应，其缺点是明显的，以致有人声称当代认知主义已经失败①、认知心理学选择了一个错误的开始。我们此处不对"失败论"和"错误开始论"进行评价，要问的是：在现代认知心理学的起点有没有其他可能的选择呢？追溯到"二战"时期，我们注意到与信息加工认知革命同时发生的吉布森主义(Gibsonism)——吉布森的生态认知心理学。

"二战"期间，吉布森(James J. Gibson, 1904—1979)的任务之一就是为美国空军重新设计机舱内的仪表盘和机场跑道信号。他在试验中寻找飞行员经常使用的视觉线索，发现紧急起飞和着陆时，飞行员是在"直接发现"信号而不是在知觉中"构造"信号的心理表征②。吉布森由此看到实验室中的认知心理学的不可靠性，开始转向对视觉现象进行生态学研究。"二战"结束后，认知革命向着信息加工理论方向发展，但在吉布森看来，这些研究趋向都是以假设的心理机制为前提或过分依赖"S—R 分析"的"病理"(pathology)心理学，于是着手"酝酿与行为主义、联想主义和 S—R 分析作根本性的决裂"。③与赫布(D. O. Hebb, 1904—1985)在美国心理学会发表演讲鼓吹"S—R 心理学"同一年，吉布森在东部心理学会发表演讲向"刺激"(stimuli)的概念发难。④他曾提出用"克分子刺激"(molar stimuli)改良这一概念，但很快发现这样做毫无价值，于是完全放弃把"刺激"作为心理学核心概念的观点。⑤他看到：生理学家可以在实验动物身上施加刺激，但是，知觉心理学家不能也不应该这样做。一个有理解力的机体是主动的、有目的的观察者，他搜寻信息，而不是被动地接受刺激。于是观察者成为心理学中的主动搜寻者，不再是被动的接受者，"刺激"的中心地位动摇了。吉布森认为，观

①② Reed, E. (1997). The cognitive revolution from an ecological point of view. In D. M. Johnson & C. Erneling (Eds.), *The future of the cognitive revolution*. Oxford University Press, Inc., p. 261—273.

③ Neisser, U. (1997). The future of cognitive science: an ecological analysis. In D. M. Johnson & C. Erneling (Eds.), *The future of the cognitive revolution*. Oxford University Press, Inc., p. 247—260.

④ Reed, E. (1997). The cognitive revolution from an ecological point of view. In D. M. Johnson & C. Erneling (Eds.), *The future of the cognitive revolution*. Oxford University Press, Inc., p. 261—273.

⑤ Gibson, J. J. (1963). *The useful dimensions of senditivity*.

察者在知觉之前，不必形成关于他看到或听到什么的假设，研究者也不必预先假定某种研究对象的内部心理结构和先天能力。他重视通过观察等发现的外部现象。

取消“刺激”概念，吉布森理论中刺激与反应的区分就不存在了，取而代之的是两种活动的区分：搜索性活动（exploratory activity）和执行性活动（performatory activity）。搜索性活动是有目的地搜索信息，它取决于观察者期望获得什么信息和从环境中能够搜索到什么信息；执行性活动是实际开展的以达到某一目标的活动，这种活动需要信息对它的调节，即要利用它能够获取的关于它与环境的信息。① 而要保持个体与环境的接触，最基本的认知过程就是觉察（perceiving）。②

吉布森的研究虽与信息加工理论同时发生在现代认知心理学的起点，但他没有像信息加工理论者那样得到更多现代科技的支持，再加上当时科学主义思潮正处于鼎盛期，生态学的认知研究不可能成为认知革命的主流。今天回头来看吉布森的工作，不必有丝毫懊恼，因为毕竟信息加工心理学积累的材料如此丰富。在批判现代认知心理学极端还原倾向的同时，也应看到吉布森的生态学理论过于简单。我们也不赞同“失败论”和“错误开始论”，符号操作范型、联结主义直至现代认知科学（cognitive science）已经并将继续取得辉煌成就。只是这种成就必须与生态学的认知研究统一起来，才能全面构成现代认知科学的框架。

吉布森的生态认知心理学取向对心理学的其他研究领域也产生了广泛影响，尤其是儿童心理学的研究。林崇德教授曾经专门发表文章阐述儿童心理学研究中的生态化运动③。林崇德认为，发展心理学中的“生态化运动”强调在现实生活中、自然条件下研究个体心理与行为，研究个体与自然、社会环境中各种因素的相互作用，从而揭示其心理发展与变化的规律。从生态学观点看，个体是在真实的自然与社会环境中成长、发展着的，个体的心理活动不可能是一个孤立的系统，心理发展要受到环境中多种因素的影响，而这些因素又相互作用、相互

① Reed, E. (1997). The cognitive revolution from an ecological point of view. In D. M. Johnson & C. Erneling (Eds.), *The future of the cognitive revolution*. Oxford University Press, Inc., p. 261—273.

② Galotti, K. M. (Ed.). *Cognitive psychology in and out of the laboratory*. Washington, Cole Publishing Company, p. 27.

③ 林崇德：《论儿童心理学与教育心理学研究的新趋势：生态化运动》，《心理发展与教育》，1990，3：144—149。

影响,是一个复杂而又完整的系统。因此,在心理发展研究中,应将被试置于现实环境,从其与社会相互作用的宏观关系中探讨其心理系统的变化。

在此强调生态学的方法论,算是对过分科学主义倾向的否定,但又不能犯"矫枉过正"的毛病,在强调生态性的整体研究、现场研究时,不能贬压其他严格的实验研究、测量研究等。人的心理是多层次的系统、心理学的研究也是多层次的系统,那么心理学的方法理所当然地也应是一个多层次的、多元的系统,不同的方法要担负不同的任务。不同的心理学家会擅长不同的方法和从事不同的研究,有着各自不同的贡献,互相认同而不是互相贬斥可能更具有建设性。

心理学的研究至少可以在三个层面上展开:第一层面是哲学考察和理论建构,第二层面是心理生活的整体的生态学考察,第三层面是结构与过程的深层次探究,不同层面关注的焦点、期望的结果、采用的研究手段、文化倾向都有所不同。其中,第一层面更多的是属于人文倾向的研究,第三层面更多的是属于科学倾向的研究,第二层面则是一种折中。不过,这些层面研究的界限越来越模糊,研究方法的交叉使用越来越频繁,比如社会心理学领域中的实验研究、自我概念的实验研究也一样可以做得很严格和很规范。单从生态性研究看,其最典型的方法是准实验,这正是一种介于实验研究与现场观察研究之间的一种研究方法和手段。掌握和使用准实验设计可以有效地提高研究的外部效度,也能有效地解决许多课题实验研究可行性较低的矛盾:降低实验条件控制的要求,在较开放的生态系统中进行准实验研究。

所谓"准实验"也即"前实验",可以理解为:还未对研究条件进行严格控制时进行的实验;或者是因为对某些额外变量无法进行控制,只好任其存在,这时进行的实验。明知存在某些额外变量却未能加以控制,这一般出现在以下几种情况:

第一,只有单一被试组,所以研究中该被试组必须接受所有处理。但是,由于无法平衡在研究进程中的时间或空间效应,存在明显的额外变量,造成研究的内部效度下降。比如,一语文教师想在自己的教学班进行一项教学实验,其教学干预必须伴随着不可逆的时间进程,对于顺序效应的平衡存在困难,此时就可以使用单组准实验设计等方法。

第二,研究中使用的多个被试组都是现成人群,由于未对被试组进行随机化分组或配对(匹配)分组,难以做到等组,组间差异成为一种明显的额外变量,造成研究的内部效度下降。比如在教学过程中,使用现有班级作为被试组进行研

究，使用的往往都不是相等组。

第三，研究中使用了实验组和控制组，但是由于实验组的特殊性，无法找到另一个相等的被试组作为控制组，这就形成了非等组，研究的内部效度下降。本章第三节中谈到的对治疗多动症的药物进行试验的例子中实验组和对照组并非相等组，所以存在明显的组间差异，这就是不等组准实验设计方法。

在上述情况下进行研究，既不像观察法那样被动，也不像实验法那样严密控制，所以都是介于二者之间的准实验研究，其外部效度和内部效度都是介于二者之间的，其研究结论往往具有较高的生态学效度。准实验设计一方面可以为因果关系的确定提供资料，另一方面又不能得到确定的因果关系，依赖的逻辑就是一种基于证据积累之后的归纳。

第二节　单组准实验设计

常用的准实验设计方法首先可以根据被试组数分为单组准实验设计和多组准实验设计。单组准实验设计的主要形式包括单组时间序列设计、相等时间取样设计；多组准实验设计的形式主要包括不等组前测后测时间序列设计、不等组前测后测设计等。本节首先介绍单组准实验设计方法。在许多研究条件下，无法对被试进行分组或施加某种严格的控制，这时就需要进行单组准实验设计。

一、单组时间序列设计

时间序列设计（time-series design），是指在引入实验处理前后都对被试或被试组进行一系列周期性观测，然后比较实验处理前序列与实验处理后序列观测值的变化趋势是否存在差异，从而评估实验处理效应的研究设计。如果研究中只有一个被试组，就叫做单组时间序列设计（single-group time-series design），其研究过程的基本模式如图 3-1 所示。

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$

图 3-1　单组时间序列设计的基本模式

图 3-1 所示的模式中，O_1、O_2… O_8 表示一系列观测或观测结果，X 表示在观测的时间序列中引入的实验处理。该设计是仅有一个实验组的单组准实验设

计，它在实验处理前后各完成一个观测序列。下面以假设的例子来说明这一研究的过程。

假如，一位小学语文教师想研究某种新的识字教学法能否提高教学效果，于是他连续地运用旧的教学法进行了四周的识字教学，并对每一周的教学效果进行检测。其操作程序：第一周教给学生 50 个生字；第二周教给学生 50 个生字，同时检测上一周的学习效果；第三周教给学生 50 个生字，同时检测第二周的学习效果；第四周继续教给学生 50 个生字，同时检测第三周的学习效果；第五周则改用新的教学方法教给学生 50 个生字，同时检测第四周的学习效果……依此类推，第五到第八周都采用新的识字教学法。如果以百分制记录检测的结果，则八周识字教学的效果记录如表 3-1。

表 3-1　识字教学法试验结果

周　次	1	2	3	4	5	6	7	8
检测分数	30	28	35	45	67	70	85	86

将表 3-1 所示的数据表示成曲线的形式，这样就能非常直观地看出引入新教学方法后的教学效果，如图 3-2。图中垂直虚线是引入新教学法的时间，即此前的四周使用传统的教学方法，此后的四周使用新的教学方法；两条斜着的虚线分别代表引入处理前后两个观测序列的回归线。

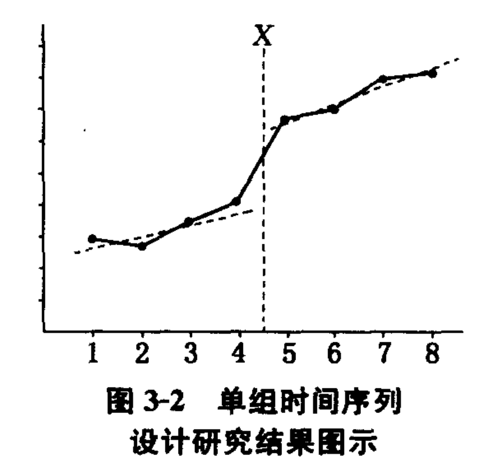


图 3-2　单组时间序列设计研究结果图示

从曲线在引入新的教学法前后的变化就可以直观地看到新教学方法的作用。这一研究就属于准实验设计中的单组时间序列设计。

时间序列设计中引入处理前后观测值的变化趋势的比较，一般从三个方面进行：

第一，观测序列回归线斜率间的比较。这一比较能够显示在引入实验处理前后的两个阶段中，被试心理或行为水平随时间变化的趋势是否一致。如图 3-2 中，两条回归线的斜率基本一致，表明新旧教学方法对每周进步幅度的影响不大。

第二，观测序列回归线截距间的比较。这种比较能够反映引入处理前后观测值总体水平的差异。如图 3-2 中，两条回归线的截距差异明显，引入处理后的

回归线截距更大，说明新的教学方法对学生学习成绩的促进效应明显。

第三，将引入处理前后的两个观测系列，作总体上的考察，看这两个阶段观测值的变化是否存在连续性。如果前后变化存在连续关系，则回归线的比较就没有意义了。

通过上述三方面的考察和比较，一般就能够检验出单组时间序列设计中是否存在实验的处理效应。当然，这一研究中存在着一些混淆变量，难以得到确定的因果关系，对此我们暂不讨论。一般来说，单组时间序列设计中的处理效应可以从回归线变化明显看出，而回归线变化的情况可概括为图 3-3 所示的几种情况。[①]

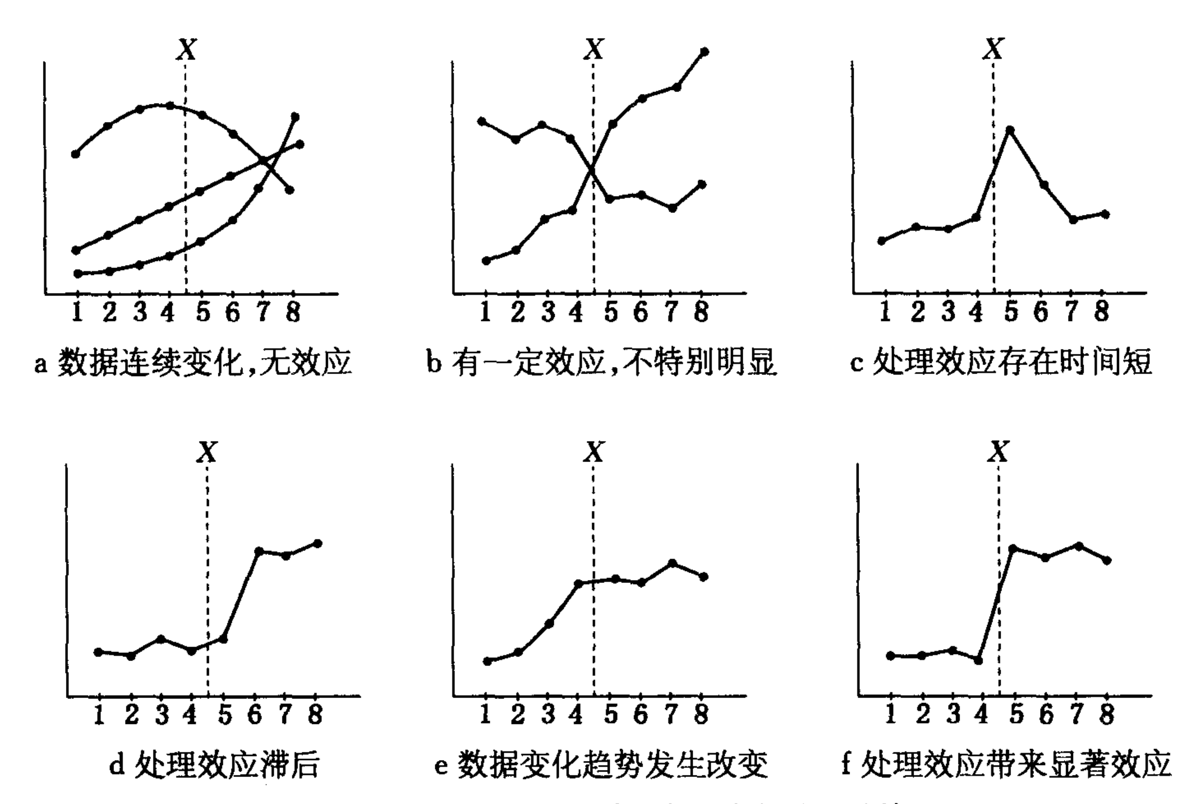


图 3-3　单组时间序列设计回归线变化的几种情况

在这些曲线中，我们能看到处理效应的不同表现：(a)引入处理前后的观测值具有连续变化关系，其变化趋势没有任何改变，实验处理没有发生明显的作用；(b)引入处理前后，观测值有一定的变化，但是变化幅度不是很大，因此处理效应不是很显著；(c)实验处理导致观测值的突然变化，而且变化幅度很大，但是随后观测值又很快回复到处理前的水平，显示出明显的处理效应，但是被试会很快适应这种实验处理，处理效应也就随之消失；(d)实验处理存在效应，但是这种

① 朱滢：《实验心理学》，北京大学出版社 2000 年版，第 36—39 页。

效应的显现有些滞后，即观测值的改变不是在施加实验处理后立即出现的，而是有所滞后；(e)实验处理改变了观测值的变化模式，处理前观测值不断上升，而且上升的速度越来越快，处理后遏制了这种变化趋势，观测值处在一个相对稳定的水平上；(f)处理效应非常明显，处理前后观测值发生了即刻的明显改变。

在许多应用研究中，单组时间序列设计可为研究者提供方便。首先，单组时间序列设计可以较好地控制"成熟"因素对研究内部效度的影响。在 $O_1 \sim O_8$ 的系列测量过程中，相邻两次测量的时间间隔基本相同，可以认为在每个时间间隔内"成熟"的发展基本相同，"成熟"仅发生在实验处理前后的可能性很小，即"成熟"仅发生在 $O_4 \sim O_5$ 之间而不发生在其他时间间隔内应是一个小概率事件。其次，单组时间序列设计可以有效地控制测验因素的干扰。由于每名被试都要接受一系列的多次测验，这就意味着是在被试适应了测验情境且测试结果趋于稳定时才施加实验处理的，这样做可以有效地减少测量偏差。此外，此种研究设计不必担心个别或少数被试中途退出实验。如果某一被试中途退出，那么他退出前的测试数据作废，这样做并不影响其他被试的实验结果。总而言之，在研究过程中，由于单组时间序列设计可以进行多次测量，研究者可以通过观察引入实验处理 X 前后测试结果的变化趋势，判断测试成绩变化与发展的趋势，进而判断实验处理 X 在时间进程中引起的某种效应。

当然，作为准实验设计，单组时间序列设计也存在不少缺点。首先，由于无控制组，因而它不能有效地识别和控制伴随实验处理发生的偶发事件的影响，不能排除那些与实验处理同时出现的附加变量的影响。因此在实验时，研究者要注意控制那些与实验处理 X 相似的无关变量，否则就会造成严重的变量混淆，无法提供充分的证据来说明实验处理的影响。其次，测验与实验处理 X 可能会发生交互作用并成为影响实验外部效度的因素，因为这种交互效应可能会歪曲或掩盖实验处理的效应大小。再次，多次实施前测往往会降低或增加被试对实验处理的敏感性，从而在被试身上产生作用而影响其实验处理后的测量成绩。最后，研究者还应注意，样本选择偏差与实验处理 X 的交互效应也可能成为影响实验外部效度的因素。由于某种原因，研究者选择的样本可能都具有某种特征，是有偏样本，不能代表样本来自的总体。这就使得样本对处理 X 发生的反应可能在样本来自的总体中并不能发生。比如，进行教法实验时，如果研究者选择的是学习能力普遍较强的班级作为实验对象，得到的结果可能就不能推广到学习能力普遍偏低的班级，研究的外部效度就比较低。

单组时间序列设计具有相对的有效性，研究者可以使用这种设计积累研究资料，为进一步严格的实验研究和科学概括提供证据。这里有三点需要说明或提请研究者注意：(1)研究中要保持实验情境的相对稳定，减少不必要的条件变化对实验结果的干扰；(2)通过单组时间序列设计实验不能得到最后的、确定性的结论，如果想得到肯定的因果关系结论，应选用有控制组参加的实验设计，或将时间序列设计运用到不同地区的不同被试组，扩大研究规模和研究范围，保证结论的概括性或普遍性；(3)由于研究中对实验条件控制不是很严格，因此研究者应充分考虑那些突发的或随意事件，详细记录研究中伴随的各种事件，这有利于对结果作出更符合实际的科学评估和解释。

二、相等时间取样设计

在实验研究中，常使用实验组和控制组的对照研究。但是当只有一个被试组时，就无法设置控制组，这时除时间序列设计外，还常常用到单组的相等时间取样设计，即对单一被试组进行无实验处理条件下的观测和有实验处理条件下的观测，而两种观测条件安排的时间取样具有一致性，其一般模式如图 3-4 所示。

$XO_1 \quad O_2 \quad XO_3 \quad O_4$

或

$O_1 \quad XO_2 \quad O_3 \quad XO_4$

图 3-4　相等时间取样设计的两种模式

在图 3-4 所示的相等时间取样设计模式中，一个被试组接受时间间距相等的四次测量，其中第一、第三次是在接受实验处理后进行的测量，第二、第四次是在未接受实验处理的常规条件下进行的测量，两种条件下进行测量的时间取样一致(分别为从第一到第三次、第二到第四次)。或者，安排的顺序相反，第一、第三次接受常规条件下的测量，第二、第四次接受处理后的测量。

比如，一位教师想研究学习反馈在学生写作学习中有无明显的作用。研究模式是：每周要求学生完成一篇作文，实验周期共四个月，其中第一个月有学习反馈、第二个月没有学习反馈、第三个月有学习反馈、第四个月没有学习反馈。有学习反馈用 XO 表示，就是每周学生完成作文后，老师都要对其作文做详细的批阅和给出恰当的评语；没有学习反馈用 O 表示，就是学生每周完成的作文交

来后，老师只是批上“阅”字和日期，不作任何修改和评价。在实验中老师悄悄地对学生的作文成绩进行成绩评定但不公布，这样就形成了如表 3-2 的实验模式。

表 3-2　写作学习中反馈效应的准实验研究模式

第一次循环	第二次循环
（有反馈的学习成绩）XO_1 （无反馈的学习成绩）O_2	（有反馈的学习成绩）XO_3 （无反馈的学习成绩）O_4

相等时间取样的测量结果可以采用类似于方差分析的方法进行处理，只不过是在对分析结果的使用上要非常谨慎才行，因为这里存在一些混淆因子，变量的效应具有某种程度的不确定性。具体地说，可对结果做三方面的检验：第一，处理条件与无处理条件间的比较，以考察存在处理效应的可能性；第二，分别在有处理条件下和无处理条件下考察时间因素的简单效应，这主要是分析研究中的时间效应或顺序效应；第三，分析实验处理与处理顺序的交互效应，以考察在时间序列中不同处理的不同效应。

这种研究设计虽然与二因素二水平的重复实验设计非常相似，但二者存在重要区别，这里的时间变量并不是两个水平，而是一个时间序列中的四个时间点或时间取样。这一设计可以较好地克服历史因素给实验结果带来的影响（即历史效应），因为实验处理间断出现，在时间取样上与无实验处理一样，而这一额外变量的影响与实验处理的变化正好保持步调一致的可能性很小。不过，毕竟这是一种准实验设计，对实验条件控制不严密，所以产生误差的因素还是很多，主要有：

第一，采用单组设计，实验处理后再重复进行做过的测验可能会增加或降低实验处理的敏感性。因为测验本身会激发被试某种心理效应，使实验结果可能不易推广到实验条件以外的情境，影响了实验的外部效度。

第二，实验安排中，实验处理的间断出现会使被试产生新异感，并暴露实验目的，由此产生实验的霍桑效应，因此实验的结果也会降低外部效度。

第三，选择偏差与实验处理的交互作用可能会降低实验的外部效度。就是说，实验采用单组被试，在选择被试时有可能选择了更适宜接受实验处理的或更不适宜接受实验处理的被试，这就出现了被试取样与实验处理之间的交互作用。

第四，实验的重复进行也会产生一系列的顺序效应。

采用单组准实验设计时要分析上述可能的变异因素，尽可能地控制这些因

素,以得到有外部效度的实验结果。或者说,只是在不适宜采用实验设计时才使用准实验设计。

第三节　多组准实验设计

在心理学研究中,为使实验结果尽量少受额外变量的干扰,在条件允许的情况下应该采用多组实验设计,即包括有实验组和控制组存在的实验设计。在准实验设计范围内,同样可以使用多个被试组以提高研究的可靠性。但是准实验中的多个被试组往往不是相等组,组间差异就是一个明显的额外变量。这里介绍两种不等组的准实验设计:不等组前测后测设计、不等组前测后测时间序列设计。

一、不等组前测后测设计

不等组前测后测设计涉及一个实验组和一个控制组,并且每组都有前测和后测,其特点是研究者不能按随机化原则和等组法选择被试,有时也不能随意地安排哪个为实验组、哪个为控制组。它的基本模式如图 3-5。

O_1　　X　　O_2
………………………
O_3　　　　　O_4

图 3-5　不相等实验组控制组前测后测设计模式

在这个模式中,O_1、O_3 表示前测,X 表示实验处理,O_2、O_4 表示后测,虚线表示两个组可能不同质,即不相等。在该设计中,研究者使用前测的目的是借助前测结果取得两个组基本相等或存在某种差异的指标,以提供两个组在控制机体变量和因变量方面最初的相等或不相等的资料,作为两个组间进行比较的基础。例如,研究者想研究不同的教材处理方法对学生学习成绩的影响。实验前,选择的教学班是两个现成班级,不可能因为实验而将两个班拆散后再随机分组,所以研究者只能随机将其中一个班作为实验组,另一个班作为控制组。依照准实验设计的模式,前测成绩是两个班(实验组和控制组)的小学升入初中的考试成绩和他们第一学期的期末考试成绩,这两个成绩作为两个班开始比较的基础。实验组实行的处理是对教材逻辑顺序的调整和增加教学过程的启发性,控制组使用传统的教学内容编排及教法。后测成绩来源于对两班学生实施的同一测验的测验分数。这就构成了一个不等组前测后测设计的准实验研究。

在这样的设计中,主要比较两个组各自的后测与前测的变化程度。变化量

的差异检验方法可以是 t 检验或中位数检验等。下面,举例说明这种研究的过程及资料分析方法。

假设某医疗研究机构研制了一种治疗儿童多动症的药物,为了试验此种药物是否真的有效,研究人员筛选了 20 名多动症儿童参加试验。为了试验的实施,他们编制了甲、乙两套学习材料,这两套材料经检验在难度等方面相当,以分别用于前测和后测。为了更可靠地进行比较,他们选取了年龄相近的某个年级一个班的学生(30 人)作为对照组。实验分三个阶段进行:第一阶段是实验组和控制组均使用甲套材料进行前测,即均在同样长的时间里学习材料甲,然后检测学习成绩;第二阶段,多动症儿童接受药物治疗,而控制组不接受;第三阶段是两个组儿童各自都学习材料乙并进行学习效果的测试,这是后测。试验的结果如表 3-3 所示。如何对表 3-3 中的数据进行定量分析呢? 这里介绍中位数检验方法。

中位数检验是通过来自两个独立总体的两个样本中位数的比较,来判断两个总体取值的平均状况是否存在显著差异。它的基本思想是假设两个总体 X 和 Y 具有相同的分布律,那么它们的取值将具有相同的平均状况,于是可以认为来自 X 的随机样本 X_1, X_2, $X_3\cdots$ X_{n1} 和来自 Y 的随机样本 Y_1, Y_2, $Y_3\cdots$ Y_{n2} 的中位数也应该大致相同。如果两个样本的中位数差异较大,则应否定两总体 X 和 Y 取值的平均状况相同的假设,或者说 X 和 Y 不具有相同的分布律。

中位数检验的具体步骤:首先将两组数据合并成一个容量为 $N=n_1+n_2$ 的样本,再找出这个样本的中位数 m。然后统计出 X 样本中大于 m 的数据个数 a,小于或等于 m 的数据个数 b; Y 样本中大于 m 的数据个数 c,小于或等于 m 的数据个数 d,有了这样的一个计数结果,就可以进行 χ^2 检验。现以上述数据为例说明这一检验过程。

第一步:计算实验组 X 的后测与前测的差异量、控制组 Y 的后测与前测的差异量:

X: 16 5 −2 30 13 15 15 12 10 15 10 20 20 5 20 20 15 5 25 14

Y: 15 −5 5 5 0 10 5 10 5 15 15 5 10 15 5 5 14 10 10 −5 15 5 10 15 5 15 5 5 20 15

第二步:计算样本 X 和样本 Y 的数据合并后数据的中位数 m:

表 3-3 不相等实验组控制组前测后测设计研究数据

实验组前测	实验组后测	控制组前测	控制组后测
20	36	40	45
25	30	55	50
40	38	35	40
20	50	60	65
30	40	65	65
40	55	50	60
30	45	35	40
20	30	40	50
50	60	55	60
30	45	50	65
30	40	40	55
25	45	35	40
30	50	30	40
40	45	40	55
50	70	50	55
30	50	60	65
40	55	60	70
30	35	50	60
20	45	55	65
50	60	65	60
		40	55
		45	50
		40	50
		30	45
		40	45
		50	65
		60	65
		65	70
		50	70
		40	55

按从小到大的顺序排列合并样本的数据：

−5 −5 −2 0 5 5 5 5 5 5 5 5 5 5 5 5 5 5 10 10 10 10 10 10 10 11 12 13 14 14 15 15 15 15 15 15 15 15 15 15 15 15 16 20 20 20 20 20 25 30

合并样本的中位数是 10.5。

第三步：统计出 X 样本和 Y 样本中大于 m 和小于等于 m 的个案数，如表 3-4所示：

表 3-4 两组成绩中位数的卡方检验用表

组　别	$>m$ 的个数	$\leqslant m$ 的个数	合　计
实验组	$a=14$	$b=6$	20
控制组	$c=10$	$d=20$	30
合　计	24	26	50

于是得到 χ^2 值：

$$\chi^2=\frac{N(|ad-bc|-N/2)}{(a+b)(c+d)(a+c)(b+d)}=5.078$$

因 $df=1$，查 χ^2 分布检验表得到临界值是 3.84，所以本研究中样本 X 和样本 Y 在前测和后测的成绩变化具有显著性差异，表明引入的实验处理对实验组产生了明显影响。从具体数据可以看出，实验组的后测成绩更明显地高于前测成绩，因此可以说，多动症儿童在服用药物之后其学习成绩提高的幅度比控制组儿童成绩提高的幅度要大。这一准实验的结果证明了该药物在治疗儿童多动症方面是有效的。

在这一准实验设计中，由于在对实验组进行前测和后测的同时，设置了一个控制组，这就可以通过对控制组中前测后测成绩的比较来观测历史的、成熟的和测验的因素的影响。当比较实验组的前测与后测之间的差异、控制组前测与后测之间的差异时，前者高于后者，这时可解释为实验处理的作用，而不能解释为由于历史、成熟和测验对实验的影响。其次，该设计基本控制了选择因素的干扰，由于两组都有前测，研究者可以通过前测结果了解被试在实验前的学习成绩和背景状况，从而使研究者得以控制选择因素。该设计模式也存在缺点：首先，研究者在选择被试时如果忽视了年龄这一因素，不同年龄的被试分别作为实验组和控制组(如 15 岁的被试为实验组，16 岁的被试为控制组)，会出现选择和成熟的交互作用，从而降低实验的内部效度。其次，由于该设计模式有前测，被试经过前测，可能对实验变量非常敏感，进而非常重视研究者的实验，造成测验的反作用。再次，研究者在选择被试时，如果出现某种偏差，选择了都具有某种特性的被试作为实验组，也可能会造成选择偏差与处理 X 的交互作用，影响实验

结果的外部效度。

二、不等组前测后测时间序列设计

在单组时间序列设计中，是对一个实验组进行一系列前测，然后施加实验处理，接着再进行一系列后测，通过比较施加处理前后测量的变化曲线确定实验处理的效应。但如果一系列的前测会使测量产生某种系列效应的积累而在量变基础上发生质变，则导致错误结论，因此可同时设置一控制组，也进行一系列的前测和后测，但不对其施加实验处理。设计模式为：

实验组：一系列前测⟶实验处理⟶一系列后测

控制组：一系列前测⟶一系列后测

设计模式中虚线以上是一单组时间序列设计，所以该设计具有时间序列设计的特点。同样，该设计还可以看作是不等组前测后测设计的扩展，即在实验处理前后各增加了三组前测和后测，所以该设计也具有不等组前测后测设计的特点。在两种设计组合的基础上，影响该设计内部效度的因素得到充分控制，即该设计控制了历史、成熟、测验、选择、选择和成熟的交互作用等因素对实验结果的影响，但测验的反作用效果，以及选择偏差与实验处理 X 的交互作用可能还会成为影响研究结果外部效度的因素。

如果将该设计应用于教学研究，由于在实验处理前后各有一系列测量，因此研究者借助这些前测、后测结果可以对一组较稳定的变化有所了解，而且也能对两组处理前后较稳定的系列变化进行比较，特别是当把这种研究纳入到教学过程时，学生一般不太了解研究者要进行的研究，这就有可能控制某些影响外部效度的因素。例如，在教学中发现，初一学生对“零”的概念，以及零不能做除数的理解相当模糊，有的学生完全靠死记硬背来处理这类问题。这时可以采用该类准实验来了解学生对有关问题的理解程度以及理解过程的变化趋势。在一所学校选择两个数学成绩较接近的班级，分别作为实验组和控制组。研究者编制 4 份试卷，其中以零的概念、有关的运算等问题为主，每隔 1 周时间对两班学生同时进行测试，将测试结果用分数记录下来。接着，对实验组进行 2 节课的辅导、练习训练，研究者再编制 4 份与前测等价的试卷，每隔 1 周时间对两班学生同时进行测验，将测验的结果用分数记录下来，分别计算两班学生各次测验的平均分

数，这样，实验组与控制组各有 8 个平均分数。下面将用统计方法描述学生对“零”的理解程度的变化趋势，以及实验组学生成绩与控制组学生成绩的比较。

可以采用两种统计方法进行检验：

方法一较为简单，类似于不等组前测后测设计。求出实验处理 X 前的 4 个平均数的共同平均数，并求出控制组的前 4 个平均数的共同平均数。接着，求出实验组和控制组的各自后 4 个平均数的共同平均数，这样，就可以采用不等组前测后测设计的统计分析方法，即将实验组的前测、后测的增益与控制组的前测、后测的增益进行差异的独立 t 检验、独立 Z 检验或中位数检验。

方法二较复杂，类似于单组时间序列设计。对实验组的 4 个前测成绩、4 个后测成绩以及控制组的 8 个成绩按单组时间序列设计的方法，求出各自的回归直线方程，这样就出现了三条回归直线方程，实验组前测成绩的回归直线方程记为 L_1、实验组后测成绩的回归直线方程记为 L_2、控制组的回归直线方程记为 L_3，按前面给出的方法检验直线 L_1 与 L_3 是否有显著差异，以确定两组在选择上是否存在偏差；检验直线 L_1 与 L_2 或者直线 L_2 与 L_3 是否存在显著差异，以确定接受实验处理所产生的效果。在实际问题中，还要看研究问题的具体情况进行显著性检验。

多组准实验设计比较好地控制了多方面的额外变量，又常使用现成的群体作为被试组，研究情境比较接近被试真实的生活、学习和工作环境，具有研究的可行性，也具有较高的外部效度。所以，这种准实验设计是实际使用较多的研究设计。

阅读材料 3-1

国内关于生态心理学的研究

国内，最早关注生态心理学(ecological psychology)的是朱智贤教授，他把生态心理学作为发展心理学的一种新取向，对它的产生及在发展心理学中的基本观点、方法和意义作了简要介绍。随后，林崇德、申继亮等人在介绍发展心理学研究现状时继承了朱老的观点。20 世纪 90 年代开始，陆续有国内学者对生态心理学进行介绍和评述，其中张利燕的《心理学研究的生态学倾向》(1990)一文对其发生的必然性、演变历程以及它在心理学发展中的意义作了较系统的介绍，不过她在介绍生态学取向的历史渊源时只提到勒温的理论。还有一些著作和文

章多是就某一领域内的生态研究取向进行介绍的，其中主要有：

著作类：主要包括申荷永的《社会心理学》，将生态心理学作为社会心理学的内容之一进行介绍；张文新的《儿童社会性发展》，介绍了生态心理学用到的一些“非实验性数据因果关系推断的方法”。

文章类：主要包括郑雪的《认知操作和认知方式与生态文化因素的关系》(1995)和《生态文化与感知觉》(1995)；邓铸的《认知研究生态学趋向及两种文化的缓和》(2000)和《文化分裂及对当代认知研究范型的反思》(2001)；贺革的《试论心理学研究方法的新进展》(2000)；左斌的《西方社区心理学的发展及述评》(2001)；管键的《生态系统心理治疗的理论述评》(2002)；肖二平的《生态心理健康》(2002)等。

另有一些文章是对与这种取向有关但已成为一门新学科的介绍：傅荣、翟宏的《行为、心理、精神生态学发展研究》(2000)；刘婷的《生态心理学研究述评》(2002)。总体来说，这些研究还比较分散，算不上是对生态心理学的系统考察。

对生态心理学的专门性和系统性研究首先是吉林大学的博士生秦晓利，他完成的博士学位论文是《面向生活世界的心理学探索——生态心理学的理论与实践》。秦晓利(2003)认为生态心理学是面向生活的心理学，是采用生态学方法论原则的心理学，并将它定义为运用心理学的视角与方法研究人与其环境关系的一门科学。他对生态心理学作了两种区分：一种是依据它们使用的方法来定义的，借用生态学的理论与方法，是理论型的生态心理学，并且把它称为生态学的生态心理学，包括巴克和吉布森等人的理论；一种是依据问题来定义的，主要是针对生态危机而言的，是应用型的生态心理学，称之为生态危机的生态心理学，包括罗杰克以及温特等人的思想和研究。他提出的生态心理学定义也正是对这两种理解的整合。他的博士学位论文主要以两种生态心理学为线索，对生态心理学的界定、背景、元理论和两种生态学取向的研究与实践进行了阐述，其理论贡献和特色在于：对两种生态心理学作了区分；对生态心理学的发生及演进背景做了较全面的分析；试图在元理论层面上对生态心理学进行整合，即在生态世界观、人性观和以心灵为研究对象以及它的方法论基础上对之进行整合；对两种生态心理学的理论和实践进行了较详细的论述；对生态心理学的贡献和局限给予了恰当的评论。

——资料来源：《生态心理学的理论审视》(易芳，2004)①

① 易芳：《生态心理学的理论审视》，南京师范大学博士学位论文，2004：3—4。

阅读材料 3-2

研究的内部效度和外部效度

前面专门讨论过因变量的效度，它是指一因变量在反映它所要反映的某种事物属性方面的可靠性，简单地说，就是它是否真的测量到它想要测量的指标。那么，对于一项研究来说，它的效度又是指什么呢？类似地，它是指一项研究是否查明了它想要查明的变量关系。这有两种表述的视域：实验情境内、实验情境外，于是就有了研究的内部效度(internal validity)和外部效度(external validity)之分。

1. 一项研究的内部效度

一项研究的内部效度是与该研究所得结果的解释唯一性有关的，也就是说，如果一项研究得到的结果有且只有一种解释，那么该研究的内部效度就高；如果一项研究得到的结果有且不止一种解释时，那么该研究的内部效度就低。因此，研究过程中所有导致对结果有其他解释的因素都被称为"对内部效度的损害"。比方说：变量 X 的变化总是一贯地、明显地伴有变量 Y 的变化。如果研究能肯定地回答变量 X 的变化是引起变量 Y 变化的直接原因，那么这个研究就具有内部效度。相反，任何其他可能引起变量 Y 变化的因素(除变量 X)都是对其内部效度的损害。

例如，要研究室温与问题解决能力之间的关系。研究中，改变室内温度，并测量每一温度条件下被试解决问题的得分，假定结果显示：当温度上升时，被试解决问题的成绩有明显下降。研究者想用温度和解决问题能力间的因果关系来解释和说明这一结果：温度升高引起解决问题能力的下降。那么，就内部效度而言，就是要看实际研究过程中除温度外，是否还有其他可能的因素引起被试问题解决能力的下降。

假设研究是这样进行的：首先将一组被试安排在室温为 20 摄氏度的房间中，然后将温度升至 30 摄氏度，再升至 35 摄氏度，每一温度下都给出问题让被试解决，并测量他们的成绩。结果显示：温度上升时成绩就下降。尽管成绩的下降可能是温度上升引起的，但也有可能是被试的疲劳引起的：他们在第一组问题中(20 摄氏度)表现不错，但到第二组(30 摄氏度)有所下降，等到呈现第三组问

题(35摄氏度)时,被试已筋疲力尽。这里观测到的成绩下降也可以用疲劳解释。这样,观测到的结果就可以有两种解释,解决问题能力的下降可能由温度引起,也有可能由疲劳引起。尽管这一实验的结果很明朗,但结果的解释却出现了不确定性,内部效度受到了损害。

现在假设研究者改用另一种方法:在每种温度下使用一个独立的被试组,以期排除疲劳因素的影响。首先将第一组20人安排在20摄氏度条件下参加实验,接着是第二组20人安排在30摄氏度条件下参加实验,最后的一组20人安排在35摄氏度条件下参加实验。结果还是显示,当温度上升时问题解决的成绩下降。然而,成绩的下降仍然可能由温度以外的其他因素引起,如被试的组间差异也能导致这一实验结果。也许前20个被试精力较为充沛,积极性较高,他们较早地被分配到实验中去,可能会感到兴奋,并想有好的表现;相反,最后参加实验的20个人很可能对研究不太感兴趣,也就没有想着要表现好,此时,观测成绩上的差异,可能是由于各组间被试的积极性差异造成的。同样,这时对结果也有两种甚至两种以上的可能解释。尽管研究者可以宣称问题解决的能力与温度有关,但这种解释的内部效度却存在疑点,即内部效度不高。

综上所述,可以看到,任何伴随自变量变化或实验处理而出现的额外变量都可能会导致研究内部效度的下降,这通常包括被试差异、实验者差异、环境差异、测量的顺序效应等。要提高研究的内部效度,就是要加强对实验中额外变量的控制,防止除自变量以外的其他任何因素成为混淆因子。

2. 一项研究的外部效度

外部效度是指在脱离研究的种种约束后,研究结果还能成立的程度。这种结果可以推广到其他人群、其他背景、其他测量吗?每一项研究都是在一特定的时间和地点进行的,有特定的被试、指导语、测量技术和实验程序,尽管研究本身具有特异性,但研究者通常期望他们的结果不是特异的,而是可以推广到研究情境之外的。

例如,斯特拉克、马丁和斯特普在1988年主持的一项研究表明,在牙齿间放一支笔(迫使其微笑)的人和在嘴唇间放一支笔(迫使其皱眉)的人相比,前者比后者更有可能认为卡通片有趣。尽管这项研究是在1988年用伊利诺斯州大学的学生做的,但研究者通常理所当然地认为这一结果可以推广到实验范围之外。就是说,如果其他人在不同的时间和地点来接受测试,应该也会得到同样的结果。

对于有些研究，如前面的微笑皱眉实验，其结果的外部效度并不存在多大争议，但也确实有很多研究存在这方面的问题，使人们对其结果的推广效果产生怀疑，常见的情况是将研究结果从严格控制的实验室推广到不加控制的现实世界。将研究结果推广到特定研究之外的效度问题，就是外部效度问题。研究中任何制约结果通用性的特征都是对外部效度的损害。研究的目的是得到能反映两变量间存在的真实关系的结果，如果研究结果能同样地适用于特定实验范围之外的其他时间、情境和人群，则认为此研究具有外部效度；相反，如果结果只在研究限定的约束条件内有效，则此研究缺乏外部效度。

因此，任何造成研究条件更具有特异性的因素都会带来对研究外部效度的损害，这一般包括被试特征、主试特征、研究的操作程序和测量方法等因素，这些因素的特异性越强，越有可能导致研究情境与人们日常生活情境的差异性，使研究情境中观测到的被试行为区别于人们日常生活情境中的行为，也就制约了研究结论的可推广性。

3. 内部效度和外部效度的平衡

任何一项研究的目的都在于使内部效度和外部效度达到最高，也就是说，每个研究者都想确保研究结果的可靠性，并且确保在研究所使用的特定的被试、情境及方法之外，这种可靠性仍然存在。然而，要设计和进行这样一个完美的研究几乎是不可能的，有些方法可以减少或排除对一种效度的影响，但同时却增加了对另一种效度的影响。因此，为了尽量提高效度，并为研究课题提供最好的答案，研究的设计和执行通常是一种充满了选择和折中的平衡行为。

为了获得较高的内部效度，研究者必须排除或减少混淆变量，严格控制实验，以防止额外变量影响实验结果。不过，对研究的控制会使研究情境带有较强的人为性，以致得到的结果在这种实验环境之外可能不成立。因此，提高内部效度会导致外部效度降低。相反，为了获得较高的外部效度，研究者通常会设计与现实世界非常相似的研究环境，这种研究的风险在于，与实验室中非常标准化的实验环境相比，现实世界往往有太多混乱的、不受控制的变量，努力提高外部效度会使额外变量进入研究，从而导致内部效度的下降。

从整体看，内部效度和外部效度存在着制约关系，如果一种效度较高，那么另一种效度势必会相对较低。在规划研究或评价别人的研究时，必须考虑这种基本关系。究竟哪一种效度相对比较重要，要注意控制哪一种影响效度的因素，往往要看研究的目的。

很明显，不同研究方法的效度的高低也都是相对而言的：(1)实验室研究法内部效度较高，外部效度较低；自然实验法内部效度较低，外部效度较高。(2)实验法内部效度较高，外部效度较低；准实验法内部效度较低，外部效度较高，但有时其内部效度和外部效度都不高。(3)单因素实验内部效度较高，外部效度较低；多因素实验内部效度较高，外部效度也较高。

因此，采用多因素的实验研究，并使研究情境尽量接近人的真实生活情境，可以较好地兼顾研究的内部效度和外部效度，这大概也是适应心理学研究生态化运动的一个折中！

——资料来源：《行为科学研究方法》(邓铸等，2005)①

建议阅读文献

1. 朱滢主编：《实验心理学》，北京大学出版社2000年版，第36—57页。
2. 孟庆茂、常建华编著：《实验心理学》，北京师范大学出版社2001年版，第33—35页。

复习思考题

1. 如何理解生态学方法论、准实验设计、时间序列设计、相等时间取样设计、不等组前测后测设计？
2. 如何理解心理学研究中的生态学方法论和生态学趋向？
3. 时间序列设计研究的一般过程如何？
4. 不等组前测后测设计研究结果如何处理？
5. 概括准实验设计的优缺点。
6. 如何评价准实验设计的内部效度和外部效度？

① ［美］弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著，邓铸、姜子云、蒋小慧等译：《行为科学研究方法》，陕西师范大学出版社2005年版，第99—109页。

第四章
心理物理学

本章内容提要

心理物理学是实验心理学中的经典内容，又获得了新发展，而且在现代心理学的基础实验研究和应用研究中都有广泛用途，所以它也属于实验心理学教学的基本内容。本章首先详细介绍使用经典心理物理学方法测量差别阈限的过程和数据处理的方法，并对其产生的误差、优缺点进行比较和分析；其次，详细介绍现代心理物理学方法——信号检测论的理论和方法，并结合实例对信号检测论的实验过程、数据分析、结果解释进行分析；最后，简单介绍三种主要心理量表的概念、特点和制作方法，以及两种重要的心物关系定律：费希纳定律和史蒂文斯定律。

心理物理学(psychophysics)是研究心理量与物理量之间函数关系的科学，包括经典心理物理学和现代心理物理学。经典心理物理学是费希纳在1860年发表的最小变化法、恒定刺激法和平均差误法的总称，是费希纳《心理物理学纲要》最核心的内容，后经修订沿用至今，主要用于感觉阈限的测定和心理量表的制订。现代心理物理学是指信号检测论(signal detection theory)，是随信息论、控制论的发展而出现的新的理论和方法。

第一节　经典心理物理学

一、感觉阈限及其解释

(一) 感觉阈限的操作定义

在普通心理学中，绝对感觉阈限(absolute sensory threshold)被定义为

“刚刚能觉察到的最小刺激量”，差别感觉阈限(differential sensory threshold)被定义为“刚刚能被觉察出的两个刺激的最小差别量”。①这里的“刚刚”如何测量呢？无法操作。我们可以把“刚刚”理解为感觉到的一个临界点，当刺激强度大于此点强度时就能被感觉到，当刺激强度小于此点强度时就不能被感觉到。可是，某一强度的客观刺激能否引起感觉的过程并不是绝对的，比如一个 50 Hz 的声音刺激，听力正常的人一般都能听到，但是也不排除偶然听不到。这就是说，一定强度的刺激可能被感觉到，也可能不被感觉到，存在一个感觉到和感觉不到的概率分布问题。一般来说，刺激强度越大被感觉到的可能性越大，刺激强度越小则被感觉到的可能性越小。

就个人接受某种刺激的情况来说，只要相关因素保持稳定，总可找到某种刺激强度，该个体感觉到和感觉不到这一刺激的可能性各占 50%。如果以此强度作用于该个体，让其报告有无感觉到刺激，重复进行 200 次，他报告“有”和“无”的次数大约会各占 100 次。不难推断，如果刺激强度增加，则报告“有”感觉的次数就会大于 50%；如果减小刺激强度，则报告“无”感觉的次数就会大于 50%。可见，通过实际的测试可以找到一个刺激值，在该刺激值上，被试感觉到和感觉不到的次数各占 50%。小于该强度的刺激则更多地被报告“无”感觉，大于该强度的刺激则更多地被报告“有”感觉。很明显，在这个刺激强度的左右，个体的感觉发生了质的变化，这个强度将刺激引起的感觉系列划分成了性质不同的两大部分：一个部分是，其刺激多数情况下能引起感觉；另一个部分是，其刺激多数情况下不能引起感觉。经过这样的分析，我们就能理解感觉阈限的真正含义，于是我们可以给出绝对感觉阈限和差别感觉阈限的操作定义：绝对感觉阈限就是被试在多次重复实验中有 50%的次数能够感觉到、50%的次数感觉不到的刺激强度；差别感觉阈限就是被试在多次重复实验中有 50%的次数能感觉到、50%的次数感觉不到的刺激的差异量。那么，感觉系统为什么会存在一个阈限，即为什么会存在一个“无感觉”的刺激区域呢？对此，不同的理论假设会作出不同的解释，这就是所谓的感觉阈限理论。

(二) 感觉阈限的理论解释

具有代表性的阈限理论主要包括经典阈限理论、神经量子理论和信号检测论。②

---

① 黄希庭：《心理学导论》，人民教育出版社 1991 年版，第 255—257 页。

② 孟庆茂、常建华：《实验心理学》，北京师范大学出版社 2001 年版，第 47—49 页。

1. 经典阈限理论的观点

经典阈限理论是上述感觉阈限操作定义的基本理论基础。一般可以认为,外部刺激作用于感觉器官中的感受器,引起一系列的神经冲动并将冲动传递到中枢神经系统,在中枢神经系统引起兴奋,即产生中枢效应。刺激引起的中枢效应的大小受到一系列内外因素的影响,主要包括外部刺激的强度、感受器的感受性、传递通道的效率和中枢系统的激活水平等。由于上述因素存在某种程度上的随机变化性,使得一个特定强度的刺激引起的中枢效应也会发生一定的随机变化,其强度变化呈现一个正态分布。被试能否感觉到刺激存在,取决于中枢效应的强度。当中枢效应的强度大于某一特定强度时,被试就会报告“有”;当中枢效应的强度小于该特定的强度时,就报告“无”。这样,在刺激系列中可以找到一个刺激强度,其引起的中枢效应分布中,大于和小于特定中枢效应强度值的概率各占 50%,这时的刺激强度就是上述定义的感觉阈限。该刺激阈限值不会是零刺激。

经典阈限理论揭示的阈限本质可以直观地表示成图 4-1 的形式。

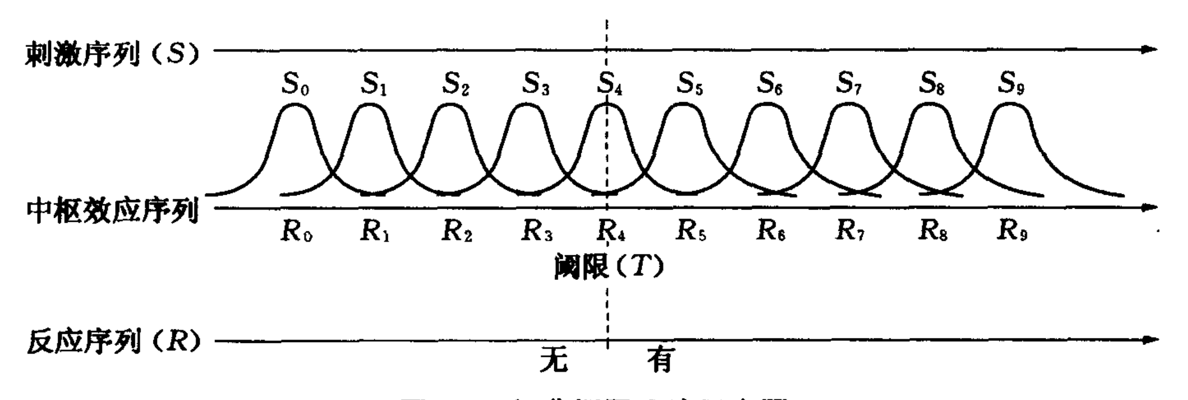


图 4-1 经典阈限理论示意图

在图 4-1 中,每一刺激均引起一正态分布的中枢效应。在 S_4 刺激条件下,引起的中枢效应分布在零刺激引起的中枢效应分布的较远的区域,其被觉察到的百分数是 50%,这时我们就说 S_4 是该被试的感觉阈限。R_4 是被试报告有无刺激的中枢效应强度标准,凡是中枢效应大于 R_4 的,被试就报告有刺激;凡是中枢效应小于 R_4 的,被试就报告无刺激。

2. 神经量子理论的观点

早在 20 世纪 40 年代,史蒂文斯(Stevens, S. S., 1941)根据响度和音高辨别实验的结果,推断基本神经过程是按照“全或无”方式进行的,即神经活

动及其功能由许多小的单元组成，每一单元都按照“全或无”方式兴奋。在某种强度刺激的作用下，一神经单元要么完全兴奋，要么完全不兴奋。当刺激强度增加时，如果增加的幅度较小，不足以使得激活的神经单元数增加，这时刺激接收者就不能觉察到刺激强度的变化；如果刺激强度增加的幅度较大，足以激活 1 个以上的神经单元（也有人认为至少要能激活 2 个神经单元），这时神经单元的激活数量就增加了，刺激接收者就可以觉察到刺激强度的变化。

需要进一步说明的是，如果原有刺激在激活若干神经单元之后还有能量的剩余，而且刺激强度的增加值与原有剩余值累加在一起可以激活一个以上的神经单元，那么这个刺激的增量也能够被觉察到。可见，在原有刺激强度不同的情况下，引起差别感觉所需要的刺激增量也不同，所以原有标准刺激强度不同时，测量到的差别感觉阈限的实验结果会有所波动。

3. 信号检测论的观点

信号检测论与经典阈限理论较为相似，也认为人类被试在接受刺激信号的作用时会在中枢系统引起一个中枢效应，而且这一中枢效应的强度在不断变化。此外，信号检测总是在噪声背景上完成的，而噪声刺激也同样会引起一个中枢效应，信号刺激和噪声刺激引起的中枢效应都是正态分布，二者有一定程度的重叠（详见本章第二节）。一般，信号刺激要比噪声刺激强些，所以二者引起的中枢效应分布也会保持一定的距离，但有部分重叠。人类被试在觉察信号时总是依据统计决策理论来进行，先确定一个中枢效应的强度标准，然后当信号刺激或噪声刺激引起的中枢效应强度小于这个标准则判断为“无信号”，当中枢效应强度大于这个标准时就判断为“有信号”。按照信号检测论，被试的判断标准就是一个阈限值，这一阈限显然是波动的，而不是一个固定的值，它使得被试能在各种噪声背景下对信号的判断达到一定的正确率。

信号检测论与前述的两个理论都有所不同。首先，它认为感觉是连续的，而不像神经量子理论所认为的是阶梯式间断的；其次，它认为阈限是一个变化的不确定的值，而不像经典阈限理论所认为的阈限是一确定的值。

上述三种感觉阈限理论虽然具有一定的合理性，但还缺乏直接的实验支持，是一个尚在探索中的问题，但即使如此，上述理论也有助于我们对感觉阈限的理解。

二、最小变化法

最小变化法(minimal-change method),也称极限法(limit method)或系列探索法等,是心理物理学的经典方法之一,它的主要特点就是刺激按照“递增”或“递减”的两种系列逐级进行微小的变化,以探测被试对刺激有无觉察的反应变化的转折点,即阈限的位置。利用这种方法既可以测定被试对刺激的绝对感觉阈限,也可以测定被试对刺激的差别感觉阈限。

(一)用最小变化法测定绝对阈限

1. 刺激系列的确定

用最小变化法测定绝对感觉阈限的实验自变量是刺激强度,其操纵方法按照递增(一般用↑标记)和递减(一般用↓标记)的两种系列变化,这两种刺激系列变化的范围和每次变化的幅度要根据仪器的可能和所测感觉通道的性质而定,一般在刺激变化的范围内确定15～20个刺激点,其中最小刺激点应保证远在阈限值之下、最大刺激点应保证远在阈限值之上。比如,根据已有研究经验知道人的听觉阈限在16 Hz左右,那么测定听觉阈限的刺激系列就可以确定在6～26 Hz之间的范围,根据刺激点数的要求,每次刺激强度的变化幅度可以控制在1 Hz。

感觉阈限的测定属于比较精细的工作,微小的误差都会对测试结果产生明显影响,所以为了实验数据的质量,需要进行更多的重复测量,一般递增和递减系列均要达到50次以上。

2. 反应记录

用最小变化法测定绝对感觉阈限的反应变量是被试的感觉,即在每次刺激条件下被试有或无感觉的变化。为记录被试的感觉变化,要求被试在每次刺激时报告“有”(感觉到刺激,一般记为＋)或“无”(感觉不到刺激,一般记为－)。在递增系列实验中,从较小的刺激开始逐渐增大,记录每次被试的反应或报告,直到被试的报告从“无”变为“有”就终止该次系列实验;在递减系列实验中,从较大的刺激开始逐渐减少,每次记录被试的反应或报告,直到被试的报告从“有”变为“无”就终止该次系列实验。在实验和反应记录过程中,还要注意以下几个方面的问题。

第一,刺激的改变是等距的。不管是递增系列还是递减系列,每次刺激强度

的增加量或减少量是固定的。

第二，刺激的起点是不确定的。每一次递增系列或递减系列实验的起点应在一定范围内随机变化，即不能每次实验均从某一刺激强度开始递增或开始递减，这样做是为了避免被试的反应定势(response set)。

第三，在被试的反应系列中，可能会出现不能确定的情况，即被试拿不准是“有”刺激还是“无”刺激，这时实际上是被试的感觉有了变化，可以中止该次实验系列，并将反应记为“?”。

第四，在递增系列和递减系列的编排上要注意克服可能的系列效应带来的误差，如练习效应和疲劳效应等。为达到克服系列效应的目的，实验可以按照“ABBA”或“BAAB”的程序来编排，即递增和递减系列按照“↑↓↓↑”或“↓↑↑↓”顺序进行。

3. 阈限值的计算

根据感觉阈限的操作定义，阈限值应是被试报告“有”反应和报告“无”反应的次数各为50%时的刺激强度。就一个递减实验系列来说，有一个被试从“有”反应转变为“无”反应的转折点，这一转折点之前的刺激被试感觉到了，只就这一次来说被试“100%”地感觉到了；在转折点之后的刺激被试没有感觉到，只就这一次来说被试“100%”地没有感觉到，于是可以认为介于这两个刺激点之间的那个中间点就是被试“50%”次感觉得到和“50%”次感觉不到的点，这一点就是该次实验得到的阈限值位置，类似地，也可以确定在递增系列实验中的阈限值位置，由此可以得到许多次实验得到的“阈限值”。当所有实验系列的“阈限值”都计算出来后求它们的平均值，就得到实验的最后结果，即绝对感觉阈限。

现在，以音高绝对阈限的测定为例，说明上述关于自变量的确定、反应记录和阈限值的确定等过程，将这些过程全部表示成表 4-1 所示的形式。[①] 按照表4-1中记录的实验结果，先计算每一次实验得到的一个“阈限值”，即表中倒数第二行的数值。最后将全部“阈限值”平均，得到音高绝对阈限为 14.8 Hz。

使用最小变化法测量绝对感觉阈限，要注意对误差的分析和控制。那么如何分析和控制实验中的误差呢？根据实验的操作过程，实验中的误差主要包括习惯误差(error of habituation)、期望误差(error of expectancy)和系列误差(serial error)，其中系列误差主要是由于随着实验进程的增加，被试出现的疲劳效应

---

① 孟庆茂、常建华：《实验心理学》，北京师范大学出版社 2001 年版，第 48—49 页。

和练习效应。这其中的习惯误差和期望误差可以通过“↓↑↑↓”或“↑↓↓↑”的实验顺序加以克服，但是疲劳效应和练习效应难以进行较为有效控制，只能通过二者相加进行平衡，因为疲劳效应和练习效应对测试结果的影响方向相反。另外，实验中采用不确定的刺激起点也是为了克服被试的反应定势带来的实验误差。那么，我们如何考察实验中各种误差因素控制的效果呢，即如何从实验结果分析实验中的各种误差效应呢？对此，我们可以做如下分析：

表 4-1　音高绝对阈限的测定

	顺序	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	系列	↓	↑	↑	↓	↓	↑	↑	↓	↓	↑	↑	↓	↓	↑	↑	↓	↓	↑	↑	↓
刺激——纯音频率 Hz	24	+								+											
	23	+			+				+	+											+
	22	+			+	+			+	+				+				+			+
	21	+			+	+			+	+				+				+			+
	20	+			+	+			+	+			+	+			+	+			+
	19	+			+	+			+	+			+	+			+	+			+
	18	+			+	+			+	+			+	+			+	+			+
	17	+			+	+			+	+	+		+	+			+	+			+
	16	+	+	+	+	+		+	+	+	—		+	+			+	+	+		+
	15	—	—	—	+	—		—	+	+	—		—	+	?	+	+	+	—	+	—
	14		—	—	—		+	—	—	?	—	+		+	—	—	+	—	—	—	
	13		—	—			—	—			—	—		—	—	—	—		—	—	
	12		—	—			—	—			—	—			—	—			—	—	
	11		—	—			—	—			—	—			—	—			—	—	
	10		—	—			—	—			—	—			—				—	—	
	9		—				—	—				—							—	—	
	8							—				—								—	
阈限值 (*T*)		15.5	15.5	15.5	14.5	15.5	13.5	15.5	14.5	14.5	16.5	13.5	15.5	13.5	14.5	14.5	13.5	14.5	15.5	14.5	15.5
		M=14.8　SD=0.84																			

第一，实验中练习效应和疲劳效应的考察。所谓练习效应(practice effect)，就是随着实验次数的增加，被试对实验任务操作的熟练性逐步提高，反应的准确性越来越好。对于感觉阈限的测定来说，如果存在练习效应，被试对实验任务及完成任务的操作就会越来越熟练，其对微弱刺激的分辨能力越来越强，这就意味着随着实验次数的增加，测得的阈限值越来越小。如果将全部实验序列按时间

均分为前后两部分，那么前半部分测得的阈限值就应该大于后半部分测得的阈限值。如果不存在练习效应，则不会出现前半部分阈限值大于后半部分的情况。

所谓疲劳效应(fatigue effect)，就是随实验次数的增加，被试的疲劳程度越来越高，从而导致实验任务的操作绩效越来越差。对于绝对感觉阈限测定的实验任务来说，如果存在疲劳效应，那么实验次数的增加也就会导致被试疲劳程度的增加，进而会导致被试感觉分辨能力的降低，测得的感觉阈限值会提高，这时将全部实验均分为前后两部分，则前半部分测得的阈限值应低于后半部分测得的阈限值，否则就不会出现这种现象。

概括来说，研究者在实验测试结束后，可以将全部实验操作序列按时间顺序均分为两部分，对前后两部分测得的阈限值进行比较并进行差异显著性检验(可以选用平均数的差异性 t 检验)，如果前半部分测得的阈限值大于后半部分测得的阈限值，那么显示出实验中存在明显的练习效应；如果前半部分测得的阈限值小于后半部分测得的阈限值，那么显示出实验中存在明显的疲劳效应。

第二，实验中期望效应和习惯效应的考察。所谓期望效应(expectancy effect)，是指在实验系列中对感觉改变的期望所造成的被试反应偏差。对于采用最小变化法测定绝对感觉阈限来说，当采用递增系列实验时，被试期望能“感觉到”的点早点到来，这种心理活动会引起被试的提前反应，导致测得的阈限值低于阈限真值；当采用递减序列实验时，被试则会期望“感觉不到”的点早点到来，这种心理活动也会引起被试的提前反应，导致测得的阈限值高于阈限真值。这就是说，如果实验中存在期望误差，它就会导致递增序列中测得的阈限值偏低、递减序列中测得的阈限值偏高，所以总体上递减序列的阈限值高于递增序列的阈限值。

所谓习惯效应(habit effect)，是指在一个实验系列中，被试连续采用某种相同的反应方式则会导致一种习惯性偏差，即导致被试的反应改变滞后。如果存在这种习惯效应，那么在递增系列实验中，被试连续“感觉不到”的反应会导致“感觉到”的反应滞后，测得的阈限值高于阈限真值；在递减系列实验中，被试连续的“感觉到”的反应会导致“感觉不到”的反应滞后，测得的阈限值低于阈限真值。这就是说，如果实验中存在习惯效应，就会导致递增系列中测得的阈限值偏高、递减系列中测得的阈限值偏低，所以从总体上递减系列的阈限值低于递增系列的阈限值。

概括来说,研究者在实验测试结束后,可以将全部实验操作序列按递增和递减分为两部分,对两部分测得的阈限值进行比较并进行差异显著性检验(可以选用平均数的差异性 t 检验),如果递增系列的阈限值大于递减系列的阈限值,那么实验中存在明显的习惯效应;如果递增系列的阈限值小于递减系列的阈限值,那么实验中存在明显的期望效应。

(二)用最小变化法测定差别阈限

1. 刺激系列的确定

利用最小变化法测定差别阈限的刺激包括一个标准刺激(standard stimulus,简称 S_t)和一个变化的比较刺激(comparison stimulus)系列,比较刺激系列的最小刺激点远在标准刺激之下,被试能明显感觉到其与标准刺激相比较弱;比较刺激的最大点远在标准刺激之上,被试能明显感觉到其与标准刺激相比较强。在比较刺激系列中包括等距的15～20个刺激点,其中也包括一个与标准刺激相等的刺激。每次实验中呈现一个标准刺激和比较刺激,要求被试对二者引起的感觉强度进行比较。根据刺激和所测试感觉通道的性质,比较刺激和标准刺激同时呈现或先后呈现。比较刺激也包括递增和递减的两种系列,其刺激起点也要有所变化。

2. 反应记录

在差别阈限的测试中,被试的反应有三类:感到比较刺激“小于”标准刺激、感到比较刺激“等于”标准刺激或“拿不准”哪个大哪个小、感到比较刺激“大于”标准刺激,在这三种情况下被试报告“小于”、“等于或拿不准”、“大于”,分别记为“－”、“＝或?”、“＋”。以时间差别阈限的测定为例,对被试反应的记录如表4-2所示。

3. 阈限的确定

在被试反应系列中,存在着一个不确定区间(IU),就是从被试不能确定比较刺激是不是小于标准刺激的下限到被试不能确定比较刺激是不是大于标准刺激的上限之间的区域,该区域的存在就是由于差别阈限的存在,所以确定该区间的下限和上限是阈限确定的关键。不确定区间的下限和上限也是一个阈限,可定义:50%次判断为“小于”标准刺激、50%次判断为“等于或拿不准”的比较刺激强度是不确定性区间的下限(T_-);50%次判断为“大于”标准刺激、50%次判断为“等于或拿不准”的比较刺激强度是不确定区间的上限(T_+)。不确定区间的中点是被试在比较判断中实际的参照点,所以叫做主观相等点(point of subjec-

tive equality，简称 *PSE*），该点有时与实际刺激的相等点不同，即存在主观相等点偏差，叫做常误（constant error，简称 *CE*），等于主观相等点强度减去标准刺激强度 S_t；不确定区间的 1/2 就是要测定的差别阈限值（*DL*）。所以，可得到下列计算公式：

$$IU = (T_{+} - T_{-}) \qquad PSE = (T_{+} + T_{-})/2$$

$$DL = IU/2 = (T_{+} - T_{-})/2 \qquad CE = PSE - S_t$$

现以时间长短估计的差别阈限测定为例说明上述过程，测试中的刺激系列、反应记录和阈限确定如表 4-2 所示。

表 4-2　时间差别阈限的测定

顺序		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
系列		↑	↓	↓	↑	↑	↓	↓	↑	↑	↓	↓	↑	↑	↓	↓	↑	↑	↓	↓	↑
比较刺激系列（单位：秒）	*ST*																				
	0.68																				
	0.64		+																		
	0.60		+								+				+						
	0.56		+	+				+			+				+				+		
	0.52		+	+	+		+	+	+	+	+				+	+		+	+	+	+
	0.48	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	0.44	+	+	+	=	+	+	+	=	=	+	+	+	+	+	+	+	=	+	+	=
	0.40	=	?	−	+	=	=	=	?	+	=	=	=	=	?	=	=	=	=	=	−
	0.36	−	−	−	−	−	−	=	−	=	−	=	−	−	−	−	−	−	−	−	−
	0.32	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
	0.28	−				−			−	−			−				−	−			
	0.24	−				−			−	−			−					−			
	0.20	−				−			−	−								−			
	0.16	−				−				−											
	0.12	−																			
上限 T_+		0.42	0.42	0.42	0.46	0.42	0.42	0.42	0.46	0.46	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.46	0.42	0.42	0.46
下限 T_-		0.38	0.38	0.42	0.38	0.38	0.38	0.34	0.38	0.34	0.38	0.34	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.42
上限平均值 $\bar{T}_+ = 0.430$　下限平均值 $\bar{T}_- = 0.378$　不确定区间 $IU = \bar{T}_+ - \bar{T}_- = 0.052$ 差别感觉阈限 $DL = IU/2 = 0.026$　主观相等点 $PSE = (\bar{T}_+ + \bar{T}_-)/2 = 0.404$ 常误 $CE = PSE - S_t = 0.004$																					

在使用最小变化法测定差别阈限的过程中，除会产生习惯、期望、疲劳、练

习等误差效应外，还可能出现因比较刺激与标准刺激在呈现时间先后、呈现空间位置的前后或左右、上下等带来的误差，所以实验中要采取相应的控制措施，减少误差带来的影响。上述误差控制的方法一般都是采用平衡方法来实现。

在采用最小变化法测定差别阈限的实验中，除要考虑并设法控制上述误差外，还要考虑到被试的主观因素可能带来的影响，比如人格因素对不确定性区间大小的影响，也即对测得的差别阈限的影响。当个体较为谨慎时，他/她在判断比较刺激与标准刺激是否有差异时，也就会比较谨慎，除非明显感到二者的差异，他/她不会作出“大于”或“小于”的反应，这样测得的不确定性区间相对较大，因此得到的差别阈限值也会比较大；相反，如果被试比较自信、冒进，就会在有轻微的差别感觉时作出“大于”或“小于”的反应，测得的不确定性区间相对较小，差别阈限值也会比较小。这样看来，使用最小变化法测定差别阈限过程中存在一个明显的误差因素，即人格因素。人格与感受性同时对被试的反应发生影响，无法分离，这是这一方法的一个明显缺陷。

三、恒定刺激法

前述最小变化法的实验程序相当繁琐，而且存在明显的期望效应和习惯效应，在数据的使用上也很不充分。为克服这些问题，可以采用恒定刺激法(constant-stimulus method)进行阈限测定。恒定刺激法也叫次数法(frequency method)、正误法(true-false method)等，其特点是：只使用从能经常感觉到的刺激到经常感觉不到的刺激之间的5～7个恒定刺激点，而且各刺激点随机呈现，不像最小变化法那样必须按照某种顺序呈现，这就减少了实验的繁琐性。相对而言，使用恒定刺激法测定感觉阈限更为准确，使用也更广泛些。

（一）用恒定刺激法测定绝对阈限

1. 自变量的确定

恒定刺激法的自变量也是刺激强度，在强度变化系列中选择5～7个强度恒定的刺激。那么如何确定这些刺激的强度呢？首先要确定刺激的范围，即根据研究经验和预先测试找到从经常感觉不到(感觉到的概率小于5%)至经常感觉得到(感觉到的概率大于95%)的刺激强度范围。范围确定后，根据范围的宽度确定5～7个等距刺激点。刺激的呈现是随机的，各种强度的刺激的呈现次数要

相等。

2. *反应记录*

用恒定刺激法测定绝对感觉阈限时，每呈现一次刺激都要求被试报告是否感觉到，即报告“有”或“无”(如果测定差别阈限，则需要被试报告“大于”、“小于”或“等于”)。该类实验同样也要重复很多次，得到每一刺激强度对应的被试报告“有”或“无”的次数。

例如，使用恒定刺激法来测定手背某一部位皮肤的两点阈。实验选定 5 个刺激点，两点阈规的两脚距离分别为 8 毫米、9 毫米、10 毫米、11 毫米和 12 毫米，每个刺激呈现 200 次，共做实验 1 000 次，刺激按随机顺序呈现。每次呈现刺激时要求被试根据感觉报告“两点”或“一点”。统计被试报告的次数，如表 4-3 所示(Woodworth & Schlosberg, 1954)。

表 4-3　用恒定刺激法测定两点阈的实验记录

刺激(毫米)	8	9	10	11	12
回答“两点”次数	2	10	58	132	186
回答“一点”次数	198	190	142	68	14
回答“两点”百分数	1	5	29	66	93

3. *阈限值的确定*

利用表 4-3 所示的实验结果计算被试的绝对感觉阈限，有多种方法可用，其中较为简明和较为常用的是 *S-P* 作图法或直线内插法、平均 *Z* 分数法。

(1) *S-P* 作图法和直线内插法

不难推断，随着刺激强度的增加，被试感觉到刺激的概率也随之提高，因此刺激强度与感觉概率之间存在一条关系曲线，即 *S-P* 曲线，这条关系曲线或其近似曲线可以通过若干实验数据得到。这一曲线经过或近似地经过“感觉到”的概率为 50%的坐标点，该点的横坐标就是感觉到和感觉不到的次数各为 50%的刺激强度，这就是绝对感觉阈限。根据表 4-3 的数据，得到两点阈实验中报告“两点”的 *S-P* 曲线，如图 4-2。从图 4-2 中纵坐标为 50%处作水平线交于 *S-P* 曲线上，再从交点处向横坐标作垂线交于横坐标轴，该交点对应的横坐标就是要测试的感觉阈限值。从图 4-2 可以看出，所求的阈限值在 10.50 稍偏向 11，可近似定为 10.60。

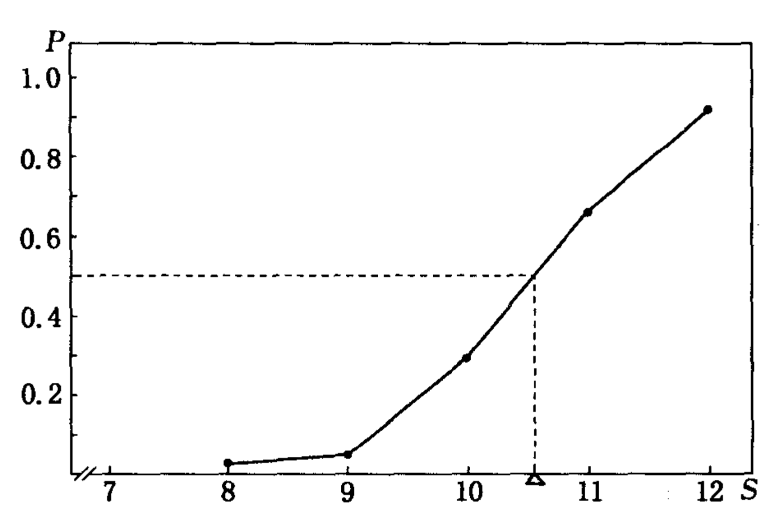


图 4-2　恒定刺激法测定感觉阈限中的 *S-P* 作图法和直线内插法

作图法是最简单的阈限值确定方法，如果实验次数很大，实验误差控制比较严密，使用标准的坐标纸来进行作图计算，可以通过 *S-P* 作图法得到较为精确的结果。如果在阈限值附近，曲线接近于直线，可以直接利用左右与阈限值相近的坐标点的坐标计算阈限值，即使用直线内插法：

根据表 4-3 和图 4-2 所示，与阈限值接近的两个坐标点是：

$$(S_1, P_1) = (10, 29)$$
$$(S_2, P_2) = (11, 66)$$

设感觉阈限值为 DL，其对应的纵坐标值为 50%，则根据直线内插法公式得到：

$$\frac{DL - S_1}{S_2 - S_1} = \frac{50 - P_1}{P_2 - P_1} \qquad 即 \qquad DL = S_1 + \frac{(S_2 - S_1)(50 - P_1)}{(P_2 - P_1)}$$

$$DL = 10 + \frac{(11 - 10)(50 - 29)}{66 - 29} = 10.57$$

S-P 作图法和直线内插法在原理上是一致的，都属于较为简单的阈限计算方法。但是这两种方法的缺点比较明显，都没有充分使用实验数据，容易受到取样误差的影响，尤其是当实验次数相对较少时，取样误差和随机误差的影响就会更加明显。下面，我们介绍一种较为繁琐的方法，这种方法更充分地使用了实验数据，因此也就可以更好地克服取样误差和随机误差。

(2) 平均 Z 分数法

按照随机事件的统计规律，随着刺激强度的增加，被试觉察到刺激的概率越来越大，是一个逐渐累加的过程，这一累加过程应该是按照正态分布的累加曲线

变化的，于是在这一实验过程中就存在相互对应的三个数据：刺激强度——感觉到该刺激的百分数——对应于该百分数的 Z 分数。也就是说，在刺激系列的某一变化过程中，被试报告“有”感觉的概率呈现正态累加。在其报告“有”感觉为50%次的位置，就正好是 $Z=0$ 的位置，此时的刺激强度就是我们要测量的阈限值。

根据 Z 分数的计算方法，我们知道刺激强度变化与 Z 分数具有线性关系：

$$S = S_0 + Z \times SD$$

公式中，S 是一刺激的强度、S_0 是刺激平均值即对应于 $Z=0$ 的刺激强度、Z 是刺激变化系列中被试报告“有”感觉概率分布的标准分数、SD 是刺激系列的标准差。于是根据实验得到的数据，我们得到如下方程组：

$$S_1 = S_0 + Z_1 \times SD \quad (1)$$

$$S_2 = S_0 + Z_2 \times SD \quad (2)$$

$$S_3 = S_0 + Z_3 \times SD \quad (3)$$

$$S_4 = S_0 + Z_4 \times SD \quad (4)$$

$$S_5 = S_0 + Z_5 \times SD \quad (5)$$

将上述等式(1)、(2)、(3)相加并除以3，(3)、(4)、(5)相加并除以3，即得到：

$$(S_1 + S_2 + S_3)/3 = S_0 + [(Z_1 + Z_2 + Z_3)/3] \times SD \quad (6)$$

$$(S_3 + S_4 + S_5)/3 = S_0 + [(Z_3 + Z_4 + Z_5)/3] \times SD \quad (7)$$

在公式(6)和(7)中，包含了刺激值平均数和标准分数平均数，这些平均数可以从实验数据计算出来，因此利用这两个等式即可计算出 S_0，该值就是我们要测定的阈限值。具体计算可表示成表4-4的形式。

表4-4　用平均 Z 分数法计算两点阈

S	$\overline{S}$	P	Z	$\overline{Z}$
8		0.01	−2.33	
9	9($\overline{S}_1$)	0.05	−1.64	−1.51($\overline{Z}_1$)
10		0.29	−0.55	
11	11($\overline{S}_2$)	0.66	+0.41	0.45($\overline{Z}_2$)
12		0.93	+1.48	

平均的刺激值 9 对应的平均 Z 分数为 -1.51、平均的刺激值 11 对应的平均 Z 分数为 0.45，代入等式(6)和(7)即可计算得到 $S_0=10.54$。这一计算结果与前述的 $S\text{-}P$ 作图法和直线内插法所得结果基本一致，应该说这一结果更精确。

(二) 用恒定刺激法测定差别阈限

使用恒定刺激法测定差别阈限的过程和结果确定具有多样性，这就使得这一部分内容的介绍显得较为复杂。其实，其基本的原理和绝对感觉阈限的测定基本一致，将前述的方法根据具体实验操作过程的变化而灵活应用就可以确定差别阈限。

可以将恒定刺激法测定差别阈限的方法分为两类反应和三类反应，而不同测定方法在阈限值的确定程序上也不完全相同。

1. **两类反应的实验**

两类反应的实验操作包括三种情况：当比较刺激等于和大于标准刺激时，要求被试不作"小于"的反应，所以被试的反应只能有两种："相等"和"大于"。如果被试报告"拿不准"或"小于"则视为"等于"，因为这时被试实际上是难以确定；当比较刺激等于和小于标准刺激时，要求被试不作"大于"的报告，所以被试的反应有两种，即"相等"和"小于"，如果被试报告"拿不准"或"大于"则视为"等于"，因为此时被试实际上是难以确定；当比较刺激扩展到从小于标准刺激到大于标准刺激的范围，被试的反应应该有三种，即"小于"、"相等"、"大于"，但是为了操作和计算的便利、为了克服其人格特征对不确定性区间的影响，可以要求被试只作两类反应，即只准其报告"大于"或"小于"，即使被试难以区分大小时也要在两种反应中选择一种他感到更有可能的倾向，所以这种方法也叫做"迫选法"(forced-choice method)。下面就两类反应的两种情况分别举例说明："大于"和"相等"反应("小于"和"相等"反应的实验过程及计算与此相似，不赘述)；"大于"和"小于"反应。

(1) "大于"和"相等"反应

如测定重量差别感觉阈限的实验：标准刺激为 200 克，比较刺激系列是 200 克、202 克、204 克、206 克、208 克、210 克，每个比较刺激与标准刺激比较 100 次，共比较 600 次。实验中被试的反应有"大于"和"相等"两种。记录每次实验中被试的反应，然后统计出各种刺激下被试报告"大于"或"等于"的次数及其百分数，结果如表 4-5 所示。

表 4-5 重量感觉的差别阈限的测定("大于"和"相等"反应)

比较刺激(克)	200	202	204	206	208	210
"大于"反应的百分数	3	29	45	73	88	96

在这种实验模式中,因为只有"相等"和"大于"两类反应,在"大于"反应的概率为 50%的刺激点实际上就是被试从能够感觉到比较刺激与标准刺激的差别过渡到感觉不到比较刺激与标准刺激差别的转折点,这一点的刺激值与标准刺激的差异量就是要测定的差别阈限,所以这一结果的计算过程与用恒定刺激法测定绝对感觉阈限的计算方法相似。可以利用 *S-P* 作图法或直线内插法,也可以利用平均 *Z* 分数法。

(2) "大于"和"小于"反应

当比较刺激系列扩展到从小于标准刺激到大于标准刺激的范围,被试感觉应是三类,即感觉到"大于"、"相等或拿不准"、"小于"。但研究者要求被试只作"大于"或"小于"的两类反应。那么,在这样的实验系列中,如何确定差别阈限呢?

可以设想,在这一实验的反应系列中,随着比较刺激强度的增加,"大于"反应的概率增大、"小于"反应的概率减少,这两条变化曲线有一个交叉点,交叉点反映被试报告"大于"和"小于"的概率相等均为 50%。在二择一反应系列中,两种反应的概率均为 50%,实际上是一个完全随机反应的概率,可以将其看作是被试对比较刺激与标准刺激的差异根本没有觉察,也就是"100%"的感觉不到刺激的差异,这是一个主观相等点。理论上讲,在实验系列中存在一个"0%"感觉到"大于"标准刺激的点,这也正好是"100%"感觉到"小于"的点,那么 25%感觉到"大于"的刺激点正好介于主观相等点与完全感觉到小于标准刺激的刺激点之间,可作为不确定性区间的下限;"100%"感觉到"大于"标准刺激的刺激点就是被试能完全感觉到"大于"的刺激点,那么 75%感觉到"大于"的比较刺激点正好介于主观相等点与完全能感觉到"大于"的刺激点之间,可作为不确定性区间的上限。简言之,就是在使用恒定刺激法测定差别感觉阈限时,如果采用迫选法,那么就以 25%感觉到"大于"的刺激点作为不确定性区间的下限、以 75%感觉到"大于"的刺激点作为不确定性区间的上限,这样得到的差别阈限也因此叫做 25%的差别阈限或 75%的差别阈限。这样,"大于"和"小于"反应系列中差别阈限的确定就转换成确定 25%报告"大于"或"小于"的刺激点和 75%报告"大于"

或“小于”的刺激点，只要采用与前述的绝对阈限确定相近似的方法确定了这两个刺激点，就可以得到差别阈限值。

2. 三类反应

迫选法毕竟在某种程度上不符合客观实际，当刺激系列扩展到大于和小于标准刺激的范围时，被试有三种反应存在，即“大于”、“相等”和“小于”，可以分别记为“＋”、“＝”、“－”。那么，如何利用这样的实验结果来确定差别阈限呢？使用这种实验测定差别感觉阈限，就是明确地承认被试在知觉判断中存在一个不确定区间，这个不确定区间的上限，应该是有50％的次数感觉到比较刺激“大于”标准刺激；不确定区间的下限，应该是有50％的次数感觉到比较刺激“小于”标准刺激。测得了不确定区间的上限和下限之后，就可以计算其差别阈限和主观相等点了。而不确定区间上限与下限的确定可以使用前述的直线内插法、平均Z分数法或作S-P图等。

但在使用恒定刺激法测定差别阈限的实验中，如果使用三类反应，被试的态度等人格因素就会对实验结果产生明显影响。我们在讨论最小变化法时已讨论过这一问题。在三类反应中，包括“大于”、“相等”、“小于”三种回答。比较自信的被试，会较少地作出“相等”的回答，其反应的不确定区间就较小，如果他根本没有“相等”的反应，则得到的不确定区间为零；比较谨慎的被试则只在具有相当把握时才会作出“大于”或“小于”的回答，所以就会作出更多“相等”的回答，其不确定区间就大，测量得到的差别阈限也就比较大。这样看来，使用三类反应测量差别阈限时，测量到的似乎不是被试的辨别能力，而是被试的态度或人格因素。因此，在使用恒定刺激法测量差别阈限时，两类反应方法更受欢迎。

四、平均差误法

平均差误法(average error method)，也称平均误差法、调整法(adjustment method)，其实验程序是：为被试提供一个标准刺激作为参照点，然后让被试自己调整一比较刺激，直到被试感觉到他自己调整的比较刺激与标准刺激相等，记录被试调整后比较刺激的实际大小与标准刺激的差异，即误差或差误值(error)，重复多次后即可得到一个平均的误差或差误，以此方法测量被试的绝对感觉阈限或差别感觉阈限，就叫做平均差误法。下面先举例说明如何使用平均差误法测量被试的差别感觉阈限。

例如，在被试正前方 2 米远处立一 1 米高的标杆，标杆的高度是固定的。然后交给被试一个装有位置可调的滑块的长杆，长杆带有长度标尺，每一次实验中要求被试调整滑块的位置直到他认为在长杆上标出的一个长度与前方标杆的高度一样时为止，这时我们就可以记录被试实际调出的长度并与前方标杆比较，求出被试每次调整的误差。如果要求被试调整很多次，其中有一半的实验是要求被试从某一段杆长明显短于标杆时开始调整，另一半实验则是要求被试从某一段杆长明显长于标杆开始调整，则可以认为被试调整误差的绝对值的平均值等于被试的差别感觉阈限(每一次调整结果与主观相等点的差值的绝对值平均数也可以算作差别阈限)、被试所有实验中调整得到的长杆的长度平均则为其主观相等点、所有误差的代数和平均则可以得到被试实验中的常误。

在使用平均差误法测定差别阈限的实验中，把标准刺激强度设为"0"时，测得的就是绝对感觉阈限。比如明度绝对阈限的测定程序为：当一个刺激的明度较大时，被试可以看到这一刺激，然后要求被试逐渐减小刺激的明度直到他看不到这一刺激停止调整；当一个刺激的明度非常小，被试在看不到这一刺激的条件下开始增加刺激的明度直到刚刚看到这一刺激停止，如此反复经多次实验得到结果的平均就是该被试的绝对感觉阈限。因为，在这一实验程序中，标准刺激强度为"0"，被试每一次调整的误差就是其调整的结果减去"0"，本身就是被试调整的结果，所以将被试多次调整的结果直接平均也就得到了绝对感觉阈限值。

看来，平均差误法的操作比较简单，而且被试参与程度比较高，积极性也比较高。

五、三种心理物理学方法的比较

费希纳的三种心理物理学方法成为实验心理学中最经典的内容，在现代实验心理学的基础研究和应用研究中仍有广泛应用。这三种方法各有特点，各有优缺点，对此可从以下几个方面来比较。①

1. 从感觉阈限的含义上来比较

最小变化法最符合感觉阈限的操作定义，同时因为被试知道刺激呈现的秩序，能把注意力集中到特别需要注意的地方，从而取得较好的实验结果。但也正

① 杨治良：《实验心理学》，浙江教育出版社 1998 年版，第 186—187 页。

是因为被试知道实验的操作秩序,则容易产生期望误差、习惯误差。这种方法适宜测定可以按小阶梯变化的刺激的感觉阈限,不适宜测试连续变化的、适应性强的刺激的感觉阈限。

2. 从被试方面来比较

最小变化法和恒定刺激法都要求较多的实验次数,被试被动参与实验,显得单调,容易造成被试的疲乏和厌烦。但是在平均差误法实验中,被试的主动参与能激发其实验兴趣。从此方面看,平均差误法优于最小变化法和恒定刺激法。

3. 从误差方面来比较

三种实验方法都会产生较大误差。极限法产生的是期望误差、习惯误差和系列误差。恒定刺激法用于那些不方便随意改变的刺激的感觉阈限较方便,但是在使用三类反应时,不确定区间受主观因素影响明显。使用两类反应虽然可以避免不确定区间稳定性方面的问题,但是迫使被试作出某种反应很不自然,容易产生逆反情绪。还有,在恒定刺激实验中,刺激随机呈现,被试具有较多猜测的成分。平均差误法实际上测到的只是一个近似值;结果不能与其他方法的结果相比较。

4. 从效率方面来比较

平均差误法的效率最高。恒定刺激法的每一组数据都可以用上,效率也较高,但是极限法实验次数多、使用数据不充分,所以效率最低。

以上各方法的优缺点都是相对而言的,使用哪一种方法要根据实验要求、被试情况和刺激性质而定。但要对两个阈限值进行比较时,这两个阈限值必须是用相同方法测试的。

第二节　信号检测论

经典心理物理学借助感觉阈限来测量感觉器官的感受性,具有一定的科学性。但在实验过程中,被试的判断明显受到其他心理特征的影响,与其掌握的主观判断标准有关。在经典心理物理学方法中,被试感觉器官的感受能力、主观判断标准等因素是交织在一起而无法分离的。那么,如何克服主观判断标准对感受能力测量的影响呢?这就要使用信号检测论方法。这里,我们将简单介绍信号检测论的数学基础和基本概念,之后重点介绍其在心理学研究中的应用。

一、信号检测论的数学基础

“二战”期间，人的信号接收与分辨能力显示出重要而特殊的意义，成为一些军事科学家特别关注的课题，再加上信息论、控制论的提出与成熟，特别是雷达技术的影响，信号检测论的基本形态逐渐形成，它概括了各种概率(probability)判断事件中的特征和规律，提出了人们在概率事件判断中的基本心理机制：存在一个信号识别的主观标准，以此标准作为信号识别的判据，人们并不企求每一次信号识别的成功，但会追求整体利益最优。所以，信号检测论的数学基础是概率判断。实验心理学教材普遍采用骰子游戏来说明这一概率判断的原理。

现在有三个骰子，其中两个是正常的，其六面分别有1、2、3、4、5、6个点子；一个是特殊的，其三个面没有点子、另三个面都有3个点子。现在，随意抛出三个骰子，报出三个骰子朝上一面点数的总和，让你猜出特殊骰子朝上一面的点数是“0”还是“3”。比如，总点数是5，你猜多少？总点数是8，你猜多少？总点数是11，你猜多少？……在这样的猜数游戏中，你能否保证每一次的回答都是正确的呢？可能性很小。但是，能否保证在若干次实验后总体上成功的次数多于失败的次数呢？这却是很容易做到的。

现在来分析，各种总点数情况下第三个骰子为“0”或“3”的相对概率：

三个骰子的总点数：	2	3	4	5	6	7	8	9	10	11	12	13	14	15
第三个骰子为“0”的相对概率：	1	2	3	4	5	6	5	4	3	2	1	0	0	0
第三个骰子为“3”的相对概率：	0	0	0	1	2	3	4	5	6	5	4	3	2	1

图4-3　各种总点数情况下第三个骰子朝上一面的点数为“0”或“3”的相对概率

从图4-3可以看出，当三个骰子朝上一面的总点数为2时，只能有一种可能的点数组合方式，即两个正常骰子朝上一面点数均为1、第三个特殊骰子朝上一面点数为0而不可能为3；当总点数为6时，存在7种可能的点数组合方式，这里第三个骰子点数为0的有5种可能、为3的有2种可能；当总点数为10时，存在9种可能的点数组合方式，这里第三个骰子点数为0的有3种可能、为3的有6种可能……总体来看，当三个骰子的总点数小于或等于8时，第三个骰子点数为“0”的概率大于为“3”的概率；当三个骰子的总点数大于或等于9时，第三个骰子点数为“3”的概率大于为“0”的概率。所以，如果报出的三个骰子总点数小于或

等于8,就猜第三个骰子的点数为“0”;如果报出的三个骰子总点数大于或等于9,就猜第三个骰子的点数为“3”,这样就可以保证成功的概率大于失败的概率。在这一骰子游戏中,猜第三个骰子是“0”或是“3”所依据的就是不同情况下事件发生的概率分布。

在现实生活和心理学实验中,类似于骰子游戏的情况很多,比如军事上依靠雷达监视领空、领海,当雷达屏幕上出现某一信号时,雷达兵就要作出判断:在雷达监控的范围内是否出现敌情,雷达捕捉到的信号是某种干扰信号或环境噪声,还是有敌方的飞机或舰艇侵入?如何判断,一般根据经验和以往计算得到的概率分布来判断,但这里总会设定一个心理标准——感觉强度或雷达检测到的信号强度。在心理学实验中,当被试有一个微弱的感觉时是判断为“有”刺激还是判断为“无”刺激呢?这与用雷达监控是相似的。

很显然,在上述信号判断实验中,有成功也有失败的可能,而每一次作何种判断,除与概率分布有关外,还与判断者的分辨能力、主观标准有关。依靠信号检测论,可以根据被试判断的结果将其分辨能力与判断标准分离出来,这是它优越于经典心理物理学方法的地方。下面我们就来介绍信号检测论的基本概念和方法。

二、信号检测论的基本参数

首先需要说明的是,信号检测论的基本假设:信号(signal)和噪声(noise)均是可能被信号检测系统检测到的刺激,而信号和噪声在信号检测系统中引起的效应强度往往不是一个确定值,而是一个具有中心趋势的随机变化值,变化呈正态分布(normal distribution),而信号检测系统对刺激作出响应的性质取决于刺激的效应强度和系统设定的判定标准。很明显,在刺激效应分布确定的情况下,判断标准的设置会影响系统响应的正确率。一般而言,系统设定的标准越高,虚报噪声的概率越小,同时漏报信号的概率也会越大;系统设定的标准越低,击中信号的概率就越大,同时其虚报噪声的概率也越大。看来,要准确地描述信号检测系统的性能,必须同时记录其对信号和噪声的响应情况。

(一)基本概率

在信号检测实验的表述中,经常使用表4-6中的概率,这些概率能有效描述信号检测系统的信号检测能力。在信号检测过程中,刺激条件有两种:一是仅有噪声、没有信号;二是在噪声背景中呈现信号。信号检测系统在这两种情形下,

对是否出现信号进行判断，判断为“有信号”，则记为“YES”；判断为“无信号”，则记为“NO”，如表 4-6 所示。

表 4-6　信号检测过程中的各种概率

	YES	NO
信号+噪声($S+N$)	击中次数 f_1 击中概率 $P(y/SN)$	漏报次数 f_2 漏报概率 $P(n/SN)$
噪声(N)	虚报次数 f_3 虚报概率 $P(y/N)$	正确否定次数 f_4 正确否定概率 $P(n/N)$

根据信号检测中信号检测系统的反应记录，可计算其击中概率(probability of hit)、漏报概率(probability of miss)、虚报概率(probability of false alarm)和正确否定概率(probability of correct rejection)。击中概率与漏报概率的和为100%、虚报概率与正确否定概率之和为 100%，所以只取其中的击中概率和虚报概率就可反映被试信号检测的情况。据表 4-6，可以计算击中概率和虚报概率：$P(y/SN) = f_1/(f_1 + f_2)$、$P(y/N) = f_3/(f_3 + f_4)$ 。

很显然，被试在信号检测实验中的击中概率和虚报概率，是由被试的分辨能力和辨别标准决定的。那么，我们能否根据其信号检测判断的结果反过来计算或推断出被试的辨别能力与判断标准呢？

（二）似然比β和辨别力指数d'

如果一被试在噪声背景上检测或识别信号，并进行有无信号的判断，就会出现上述四种反应，即击中、漏报、虚报和正确否定。可以设想：如果该被试的辨别能力较强，就能更准确地判断有无信号，击中概率就较大、虚报概率较小；如果该被试的辨别能力较弱，则击中概率就小、虚报概率就大。因此，击中概率和虚报概率能反映被试信号辨别能力的大小。

如果被试在进行有无信号判断时比较谨慎，只有当感觉强度比较大时才报告有信号，那么被试对信号的漏报就会比较多，相应的虚报概率就比较小；如果被试比较冒进，有比较微弱的感觉强度就说有信号，那么，被试的漏报概率就会下降，虚报概率则会上升。由此可见，被试的击中概率和虚报概率也可以反映被试的判断标准的高低。也就是说，我们可以从被试的击中概率和虚报概率来估计被试的判断标准和辨别能力。

被试对噪声的感觉强度呈正态分布、对信号的感觉强度也呈正态分布，如图

4-4 所示。

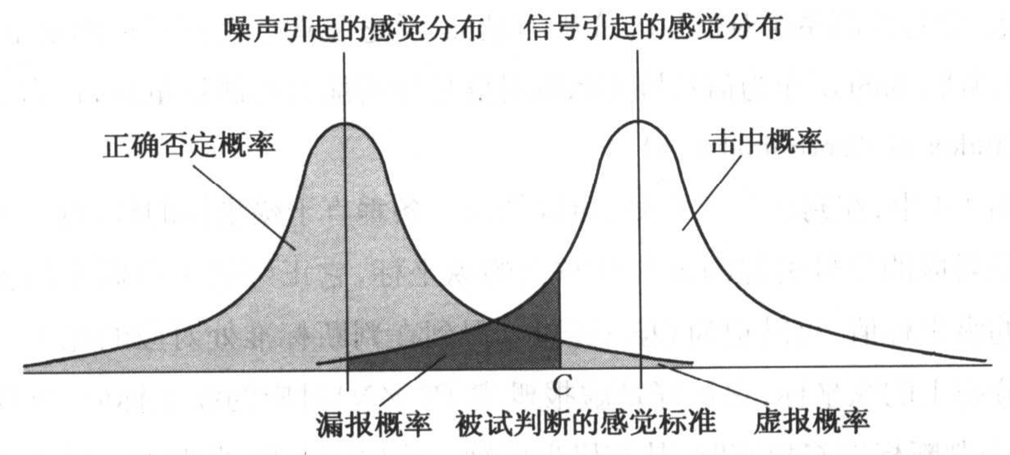


图 4-4　信号检测实验中噪声与信号引起的感觉分布及各概率分布

从图 4-4 可以看出:(1)信号引起的感觉强度是一正态分布,噪声引起的感觉强度也是一正态分布,二者有一定的重叠,这是被试判断无法避免错误的原因所在。(2)被试在进行信号识别时总会有一个标准,当感觉强度大于这个标准时就报告有信号,如果感觉强度没有达到这样的标准就报告无信号。提高判断标准时,击中概率就会减少,虚报概率也会减少;当降低判断标准时,击中概率会增加,虚报概率也会增加。所以击中概率和虚报概率是相互制约的,究竟把判断限定在什么标准上,主要取决于被试的价值取向、性格特征和外在奖惩条件(利害关系)。(3)噪声感觉分布与信号感觉分布重叠得越多,距离越小,越难区分;相距越远则越容易辨别,因此,可以两个正态分布的距离来作为信号检测者对信号的辨别力指数。那么,如何计算信号检测者的信号辨别力指数并估计被试判断标准的高低呢?

参照图 4-4,如果被试的判断标准为 C,意味着不管是信号刺激,还是噪声刺激,只要刺激引起的效应强度达到了 C 的水平,则判断为有信号,否则就判断为无信号。所以,只要在检测实验中计算得到了击中概率和虚报概率,那么就可以根据正态分布下的 PZO 转换表计算出 C 点到两个分布峰值的距离(以 Z 分数表示)。在信号引起的效应强度分布中,坐标 C 到其峰值的距离等于 Z_{SN},Z_{SN} 是以击中概率借助 PZO 转换表得到的标准分值;在噪声引起的效应强度分布中,坐标 C 到其峰值的距离等于$(-Z_N)$,Z_N 是以虚报概率借助 PZO 转换表得到的标准分值,所以两个分布的峰值的距离为:

$$d' = Z_{SN} - Z_N$$

上文已说明,信号和噪声在信号检测系统中引起的效应强度分布之峰值距离越大,信号检测系统对信号的分辨力越强,反之越弱。在信号检测论中,就以上式计算得到的 d' 作为信号检测系统对信号分辨能力的测量量数,称为辨别力指数(index of discriminability)。

图 4-4 中,在判断标准 C 处,可以做出一条垂直于横坐标的线,得到在判断标准处对应的信号引起的效应分布上的纵坐标,它正好是击中概率 $P(y/SN)$ 对应的纵坐标值,可以记为 O_{SN};还可以得到在判断标准处对应的噪声引起的效应分布上的纵坐标,它正好是虚报概率 $P(y/N)$ 对应的纵坐标值,可以记为 O_N。当判断标准 C 提高时,其在横坐标轴上的位置右移,此时 O_{SN} 增大,O_N 减小,二者的比值增大;反之,当判断标准 C 降低时,其在横坐标轴上的位置左移,此时 O_{SN} 减少,O_N 增大,二者的比值减小。看来,O_{SN}/O_N 比值可以反映信号检测过程中判断标准的高低,信号检测论将其称为似然比(likelihood ratio),记为 β:

$$\beta = O_{SN}/O_N$$

辨别力指数 d' 和似然比 β 是信号检测实验的两个基本参数,也是描述信号检测系统的两个最有效参数。辨别力指数 d' 能够直接反映信号检测系统对信号和噪声的区分能力,在具体情境中它又有不同的含义,比如在再认实验中,辨别力指数的大小反映了再认者对学习项目和无关项目的区分能力,进而反映了他对学习项目的保持程度,借此可以测量人的记忆能力;在感觉器官感受性的测量中,让被试对微弱刺激进行检测,其辨别力指数反映了被试对微弱刺激的辨认能力,也就进一步反映了被试对刺激的感受能力或感受性。似然比 β 可以反映被试在判断信号时的主观标准,它独立于辨别力指数。信号检测论的最大优点,就是将被试对信号辨别能力的测量与其判断的主观标准分离开来。

(三) 信号接收者操作特征曲线

在通常情况下,信号检测系统对信号的分辨能力是一个稳定参数,但其判别标准则会随同某些情境条件的变化而发生变化,击中概率和虚报概率也会随着情境条件、主观判断标准的变化而发生变化。一般引起判断标准发生变化的因素主要有两个方面:一个是信号出现的先定概率(prior probability),即实验之前被试了解到的信号出现的相对概率;另一个是正确反应(击中、正确否定)和错误

反应(漏报和虚报)带来的不良后果(利害关系)。研究者在长期研究中,概括出了一个估计信号检测判断标准的经验公式[1],即:

$$\beta = \frac{P(N)}{P(SN)} \times \frac{V_{n.N} + V_{y.N}}{V_{y.SN} + V_{n.SN}}$$

经验公式中,$V_{n.N}$表示正确否定奖励数、$V_{y.N}$表示虚报处罚数、$V_{y.SN}$表示击中奖励数、$V_{n.SN}$表示漏报处罚数

从公式可以看出,信号检测者信号判断的标准(β)与信号的先定概率成反比关系,即信号出现的概率越大,判断的标准越低;反之越高。同时也可以看到,信号存在时击中得到的奖励越多、漏报受到的惩罚越大,判断标准越低;反之越高。没有信号时正确否定的奖励和虚报的惩罚数量越大,则判断标准越高;反之越低。

如果改变信号检测条件,包括改变先定概率和奖惩条件,就会导致信号检测者的判断标准的改变。一般情况下,信号检测者的信号分辨能力相对稳定,这就意味着条件改变可以导致其判断标准(似然比 β)的改变,但不会导致辨别力指数(d')的改变,于是实现了判断标准与辨别能力测量的分离,这就是信号检测论方法与经典心理物理学方法相比的最大优势。

具体地说,我们可以假设在各种先定概率或各种奖惩条件下对某一被试进行信号检测能力的测试,就可以得到多组信号检测实验的记录,如表 4-7 和表 4-8 所示[2]。表 4-7 所示的实验结果显示,在先定概率发生变化的情况下,被试的击中概率和虚报概率都发生了变化,计算出来的似然比 β 也发生了变化,但是辨别力指数 d' 却保持稳定;表 4-8 所示的实验结果则显示,先定概率保持不变,但奖惩条件发生变化时,被试的击中概率和虚报概率也会发生变化,计算出来的似然比 β 也发生了变化,但辨别力指数 d' 也是保持相对的稳定。看来,实验条件的改变并没有影响对该被试稳定的辨别能力的测试。如果以不同信号检测条件下记录的被试的虚报概率为横坐标、击中概率为纵坐标,则可以得到一条二维曲线,该曲线上的各点反映了被试在不同判断标准时的击中概率和虚报概率,但曲线上各点的辨别力指数是相等的,所以这一曲线被称为信号接收者操作特征曲线(receiver operating characteristic curve,简称 *ROC* 曲线)。根据表 4-7 和表 4-8 中的实验结果可以作出两条 *ROC* 曲线,如图 4-5所示。

---

① 杨治良:《实验心理学》,浙江教育出版社 1998 年版,第 235 页。

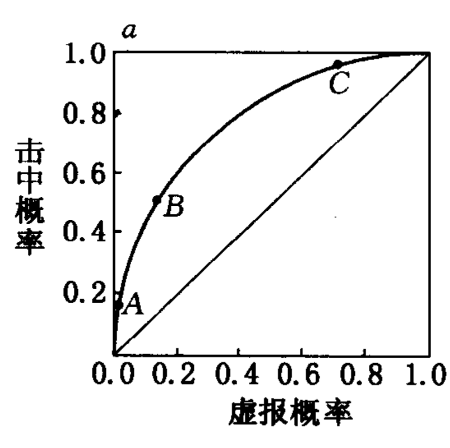
② 朱滢:《实验心理学》,北京大学出版社 2000 年版,第 94—97 页。

表 4-7　似然比 β 与先定概率的关系

各次实验	先定概率		YES	NO	计算的结果
A	0.20	SN	0.16	0.84	$\beta=4.481$　$d'=1.0$
		N	0.02	0.98	
B	0.50	SN	0.50	0.50	$\beta=1.649$　$d'=1.0$
		N	0.16	0.84	
C	0.80	SN	0.93	0.07	$\beta=0.369$　$d'=1.0$
		N	0.69	0.31	

表 4-8　似然比 β 与奖惩条件的关系(先定概率 $P_{SN}=0.50$)

各次实验	奖惩条件				YES	NO	计算的结果
		YES	NO				
D	SN	+1	−1	SN	0.31	0.69	$\beta=4.481$
	N	−2	+2	N	0.07	0.93	$d'=1.0$
E	SN	+1	−1	SN	0.69	0.31	$\beta=1.000$
	N	−1	+1	N	0.31	0.69	$d'=1.0$
F	SN	+3	−3	SN	0.98	0.02	$\beta=0.223$
	N	−1	+1	N	0.84	0.16	$d'=1.0$



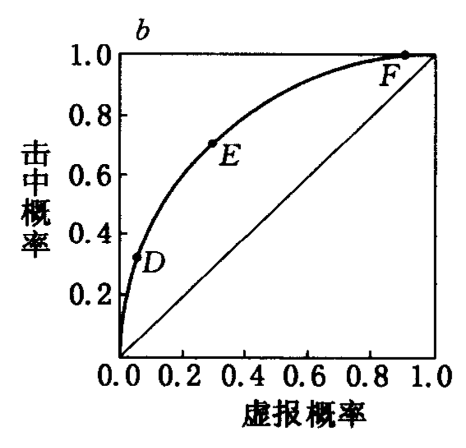


图 4-5　根据表 4-7(左边的曲线)和表 4-8(右边的曲线)所作的 *ROC* 曲线

在图 4-5 的曲线 a 上,A 点显示的是被试在先定概率最小的情况下的信号检测特征,此时因为信号出现的概率本身很低,所以被试不轻易作出“YES”反应,击中概率和虚报概率都比较低,判断标准比较高;相反,C 点显示的是被试在先定概率最大的情况下的信号检测特征,此时因为信号出现的概率本身很高,所

以被试倾向于多作"YES"反应，击中概率和虚报概率都高，判断标准比较低。很明显，被试的判别标准与信号的先定概率成反比。

图 4-5 的曲线 b 上，D 点显示的是当虚报的惩罚数相对较大时，被试会提高判断标准而尽量减少虚报，不轻易作出"YES"反应，击中概率和虚报概率都最低；F 点显示的是当击中的奖励数相对较大时，被试会降低判断标准，尽量增加击中数，倾向于多作"YES"反应，击中概率和虚报概率都最大。

此外，被试信号检测的判断标准也与其自身的性格特征有关。一般而言，做事比较谨慎的被试不会轻易作出"YES"反应，除非感觉强度比较大时，所以这类人的判断标准会比较高；相反，性格较随意或较冒进的人，则更容易作"YES"反应，其判断标准较低。

击中概率、虚报概率、判断标准和辨别力指数的关系可以表示成图 4-6 所示的形式。①

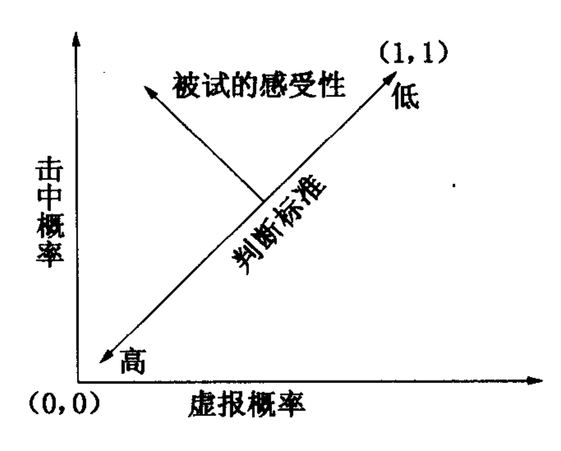


图 4-6　*ROC* 曲线特征图

从图 4-6 可知：(1)通过坐标系原点(0, 0)和坐标点(1, 1)的直线，虚报概率和击中概率总是相等的，显示出被试对信号无任何分辨能力，被试采用的回答策略就是根据先定概率而随意按比例回答"YES"或"NO"，其中有两种极端情形：当被试不管感觉强度如何，都回答"YES"，就会出现击中概率和虚报概率都等于 100% 的情形，在 *ROC* 曲线坐标系中为点(1, 1)；当被试不管感觉强度如何，都回答"NO"，就会出现击中概率和虚报概率都等于 0 的情形，在 *ROC* 曲线坐标中的原点(0, 0)；(2)被试反应的击中概率和虚报概率均较小时，其在特征曲线上的坐标点靠近左下角的原点，这时显示其判断标准比较高；被试反应的击中概率和虚报概率均较大时，其在特征曲线上的坐标点靠近右上角的(1, 1)点，这时显示其判断标准比较低；(3)如果一个被试的 *ROC* 曲线越靠近坐标系中的对角线，显示出被试的辨别力指数越低，越趋近无信号分辨能力；*ROC* 曲线越偏向左上角方向，越远离对角线，显示出被试的辨别力指数越大，信号分辨能力越强。

① 朱滢：《实验心理学》，北京大学出版社 2000 年版，第 96—97 页。

三、信号检测实验

利用信号检测论进行心理学研究的实验方法主要包括有无法(yes-no method)和评价法(confidence rating method),这两种方法既有联系又有区别,评价法可看作是对有无法的扩展。

(一) 有无法实验

这是信号检测实验的基本形式,一般的操作程序是:在一定的噪声背景上呈现或不呈现信号刺激,每次实验中都要求被试作出"有"或"无"信号的判断,记录其反应正误,进而确定其击中概率和虚报概率,查正态分布下的 *PZO* 转换表并将相应数值代入公式计算被试的辨别力指数 d' 和似然比 β,借此实验结果就可以描述被试对信号刺激的分辨能力或感受能力及其判断标准的高低。在心理学及相关的应用领域中,有无法实验可被应用于不同的具体情境,如记忆的再认实验、感官感受性的测定、产品质量检验等等。我们现在举例说明,这一实验方法的操作过程以及实验结果的处理方法。

比如,采用再认法对甲乙两人的记忆能力进行测试和比较。其实验过程是:甲乙两人在相同时间内同时学习并记忆 50 个单词,然后将这 50 个单词与另外 50 个未学习的单词混淆在一起,再逐个地分别呈现给甲乙两人,要求其回答呈现的这个词"是"或"不是"前面学习过的单词。于是得到如表 4-9 所示的测试结果,那么按照测试的结果,甲乙两人中谁的记忆能力比较强、谁的判断标准比较高呢?

表 4-9　假想的有无法实验结果

被试	被试的反应结果		
	项目	"是"	"不是"
甲	学习过(*SN*)	45	5
	未学过(*N*)	15	35
乙	学习过(*SN*)	30	20
	未学过(*N*)	10	40

这一实验结果的处理过程包括以下三个主要步骤:

第一步:计算被试的击中概率和虚报概率

甲的击中概率 $P_{甲}(Y/SN)=45/50=0.90$　虚报概率 $P_{甲}(Y/N)=15/50=$

0.30

乙的击中概率 $P_{乙}(Y/SN)=30/50=0.60$　虚报概率 $P_{乙}(Y/N)=10/50=0.20$

第二步：查正态分布下的 PZO 转换表得到相应的 Z_{SN}、Z_N、O_{SN} 和 O_N

被试甲：$Z_{SN甲}=1.28$　$Z_{N甲}=-0.52$　$O_{SN甲}=0.176$　$O_{N甲}=0.348$

被试乙：$Z_{SN乙}=0.25$　$Z_{N乙}=-0.84$　$O_{SN乙}=0.387$　$O_{N乙}=0.280$

第三步：将第二步计算的结果代入公式计算辨别力指数和似然比

被试甲：$d'_{甲}=Z_{SN甲}-Z_{N甲}=1.28-(-0.52)=1.80$

$\beta_{甲}=O_{SN甲}/O_{N甲}=0.176/0.348=0.506$

被试乙：$d'_{乙}=Z_{SN乙}-Z_{N乙}=0.25-(-0.84)=1.09$

$\beta_{乙}=O_{SN乙}/O_{N乙}=0.387/0.280=1.38$

第四步：根据计算得到的辨别力指数 d'、似然比 β，对相关问题作出具体说明

根据前述计算的结果，甲对学习项目的辨别力指数大于乙的辨别力指数，说明甲在相同条件下对项目的学习效果好于乙，因此可以认为甲的记忆力好于乙；乙的似然比 $\beta_{乙}$ 大于甲的似然比 $\beta_{甲}$，说明乙的判断标准高于甲，相对而言，乙在作出“是”学过的判断时更谨慎。

（二）评价法实验

有无法实验，要求被试根据自己的感觉判断是否有信号出现。被试在实际操作上将感觉连续体分为两个部分，即在某一标准之上报告为“有”信号，在某一标准之下报告为“无”信号。然而，实际操作中，被试的判断存在信心水平的不同：感觉强度越大，报告“有”信号的信心水平越高；感觉强度越小，报告“无”信号的信心水平越高，这种信心水平的变化在有无法实验中难以显现。于是信号检测论衍生出评价法实验，评价法实验是对有无法实验的扩展，即在有无法实验的过程中，要求被试在报告有无信号的同时评定判断的把握度，即信心指数，这时的实验方法就叫评价法。如当呈现刺激时，被试报告“有”，给出的信心指数是80%，那就等于说，被试认为此时有信号的可能性达到80%。也可以要求被试在报告有信号时给出把握等级，比如“非常有把握”的等级为5，“很没有把握”的等级为1，两者之间就有比较有把握、中等把握、较没有把握。这样，多次实验就可以得到被试在各种判断把握度上的击中概率和虚报概率，从而评估被试在各种不同判断标准上的信号检测特征。

借用朱滢先生《实验心理学》①中的假想数据来说明评价法实验的原理与过程。假设在一个信号检测实验中，SN 和 N 各呈现 600 次，在总共 1 200 次实验中，每一次都要求被试报告其在何种程度上认为有信号或无信号出现，记录其报告结果并将其报告结果汇总，如表 4-10。

表 4-10　一个假想的评价法实验结果

确信水平	1	2	3	4	5	6
有信号时各评价等级数	44	59	66	101	154	176
无信号时各评价等级数	198	162	96	66	54	24
有信号时各评价等级概率	0.01	0.10	0.11	0.17	0.26	0.29
无信号时各评价等级概率	0.33	0.27	0.16	0.11	0.09	0.04

↑ C_1　↑ C_2　↑ C_3　↑ C_4　↑ C_5

（此假想数据来自朱滢《实验心理学》的第 100 页，但在数据编排顺序上进行了调整）

附注：表中的确信水平 1、2、3、4、5、6 分别代表的确信程度是：1——认为信号出现的可能性为 0%（即非常肯定地认为没有信号出现）；2——认为信号出现的可能性为 20%（即肯定地认为没有信号出现）；3——认为信号出现的可能性为 40%（即认为可能没有信号出现）；4——认为信号出现的可能性为 60%（即认为可能有信号出现）；5——认为信号出现的可能性为 80%（即肯定地认为有信号出现）；6——认为信号出现的可能性为 100%（即非常肯定地认为有信号出现）。

根据表 4-10 中的数据就可以确定，如果被试按照不同的标准来判断“有”或“无”信号，他就可以得到各自不同的击中概率和虚报概率。比如，如果被试采用的标准为 C_1，那么他只有当能非常肯定地认为没有信号时才会报告“无”，否则就会报告“有”，这时其击中数就是 556，虚报数就是 402，击中率和虚报率都非常高，判断标准很低；如果被试采用的标准为 C_5，那么他只有当能非常肯定地认为有信号时才会报告“有”，否则就会报告“无”，这时其击中数就是 176，虚报数就是 24，击中率和虚报率都很低，判断标准很高……依此类推，就可以得到表 4-11 中所示的被试在各种不同判断标准下的击中概率和虚报概率，进而可以计算各种不同标准下信号检测的似然比 β 和辨别力指数 d'。

从表 4-11 所示的计算结果显示，被试如按照不同的判断标准来对信号进行检测判断，那么他的击中概率和虚报概率都不相同，得到的似然比不同，似然比

① 朱滢：《实验心理学》，北京大学出版社 2000 年版，第 100—101 页。

的变化正好反映了判断标准的变化。不过在各种判断标准上得到的辨别力指数是相对稳定的,即可以在误差范围内认为其未发生变化。这一结果再一次表明信号检测论可以实现辨别能力测量与判断标准测量的分离。

表 4-11 被试在各种判断标准下的信号检测特征

判断标准	C_1	C_2	C_3	C_4	C_5
击中概率 $P_{(Y/SN)}$	0.93	0.83	0.72	0.55	0.29
虚报概率 $P_{(Y/N)}$	0.67	0.40	0.24	0.13	0.04
击中概率对应 Z 分数 Z_{SN}	1.48	0.95	0.58	0.13	−0.55
虚报概率对应 Z 分数 Z_N	0.44	−0.25	−0.71	−1.13	−1.75
击中概率对应纵坐标 O_{SN}	0.133 4	0.254 1	0.337 2	0.395 6	0.342 9
虚报概率对应纵坐标 O_N	0.362 1	0.386 7	0.310 1	0.210 7	0.086 3
似然比 β	0.37	0.66	1.09	1.87	3.97
辨别力指数 d'	1.04	1.20	1.29	1.26	1.20

与有无法相比,评价法有两个明显的优势:一是信息量大,二是效率高。有无法只要求被试做有无信号的判断,被试就只能以某一标准来判断,但是实际上不同时刻被试在判断有无时的信心水平是不同的。在评价法实验中,被试不同的信心水平可以得到充分表现,所以评价法实验能够更充分地显示被试信号检测中的心理活动特征。在效率方面,一个评价法的实验过程可以相当于几次有无法的实验成效,效率大大地提高了,比如被试对信号判断的确信等级分为 6 等时,其得到的实验结果相当于 5 个有无法实验得到的结果。

最后,综合第一、二节内容,可以看到与经典心理物理学方法相比,信号检测论具有独特的价值和优势。信号检测论引入心理实验,是心理物理学的一个重大突破[①]:第一,在感受性的测量上,把主观态度与辨别力分开,获得成功。第二,在解决辨别力指数问题上,明确地建立起反映噪声变量的虚报概念,这一点对于辨别力的评估不可缺少,而且对于被试主观态度或反应偏好的测量具有重要意义。这里是说,如果不能建立噪声变量的虚报概念,只有击中概率就不能有效估计被试的似然比和辨别力指数。第三,信号检测论不仅把信号当信息,也把噪声当信息,这相对于传统心理物理学来说是一个发展,因为噪声引起的感觉分布是信号检测的主要困难,只有充分地分析噪声信息才能更好地研究被试的辨别力的性质。

① 孟庆茂、常建华:《实验心理学》,北京师范大学出版社 2001 年版,第 107 页。

第三节　心理量表与心物关系

前述的两种心理物理学方法可用于对感受性的测量，但研究物理量与心理量的关系，仅仅测量感受性是不够的，还需要了解阈上刺激与感觉量的关系。因物理量与感觉量不是简单的线性关系，不能直接使用刺激量的变化来推算心理量的变化，需要使用心理量表法来建立阈上刺激与心理量的关系定律。

所谓心理量表(psychological scale)，就是刺激的物理强度与其引起的心理强度的对照表。有了这一对照表，就可以从刺激量大小了解相应的心理量大小，也可以从感觉量大小了解刺激量大小。当然，进行比较，常常存在不同的精度要求。如只比较谁大谁小，这就是排序问题；需要比较差异量的多少，则需要度量单位，这就要把感觉强度分出衡量单位；需要比较两个刺激引起的感觉的倍数关系，必须知道二者的绝对大小，则需要一绝对零点，在这种情况下，才能衡量二者的比例关系。由此分析，心理量表包括三个不同水平，即顺序量表(ordinal scale)、等距量表(equal interval scale)和比例量表(ratio scale)。下面，我们介绍各种心理量表的定义及制订方法，同时讨论两种有代表性的心物关系定律。

一、顺序量表及其制作

顺序量表就是根据引起感觉的性质差异或强度大小将刺激排序，构成一个物理量和心理量的对照表，这种量表既无相等的单位又无绝对零点。其制订方法有两种：等级排列法(rank-order method)、对偶比较法(method of paired comparison)。等级排列法就是把许多刺激同时呈现给多名被试，让每一位被试根据自己的某种心理感受将这些刺激排序，然后将所有被试评定的等级进行平均，即得到一个顺序量表。对偶比较法，则是把所有刺激两两比较，每一次比较都得到一个优势刺激，然后统计各个刺激在全部比较中获得优势次数的多少，就可以将它们的顺序排列出来。

比如，采用等级排列法对 7 幅绘画作品进行评估，就可以得到人们对这些作品喜好程度的排列顺序。具体方法是：首先选取 n 个被试组成样本；然后要求每一被试独立地对 7 幅作品进行排序，如表 4-12 所示；最后计算各作品的平均等级，根据计算结果就可以得到相应人群对这些作品喜好度的顺序量表，如

表4-13所示。

表 4-12　5 名被试对 7 幅绘画作品的评价等级

被试号	不同颜色排列顺序(从最喜欢到最不喜欢)						
	7	6	5	4	3	2	1
1	C	E	D	B	A	G	F
2	A	C	E	D	B	G	F
3	C	G	E	A	D	B	F
4	E	C	G	D	A	F	B
5	D	C	E	G	B	A	F

表 4-13　7 幅被评价绘画作品的平均等级

被试号	A	B	C	D	E	F	G
1	3	4	7	5	6	1	2
2	7	3	6	4	5	1	2
3	4	2	7	3	5	1	6
4	3	1	6	4	7	2	5
5	2	3	6	7	5	1	4
平均值	3.8	2.6	6.4	4.6	5.6	1.2	3.8

从最喜欢到最不喜欢的绘画作品依次是:C、E、D、A/G、B、F,这就形成了一个作品喜好度顺序量表。

顺序量表制作方法可用于对客观刺激引起的心理效应的强弱进行评估,也可以用于对个体差异进行行为评定,比如研究一个班级中的人际关系、对一个管理机构中各位管理者的管理能力与绩效的评估等。现在,我们假想一个公司的人事部要对公司的 6 名中层领导进行管理能力与管理绩效进行评估,组织了 5 名评估专家,采用的方法是对偶比较法,即每一位评估者都对 6 名被评估者作两两比较,对每次被比较的两人进行评估,管理能力和管理绩效强者记为 1 分,弱者记为 0 分,结果如表 4-14 所示。表 4-14 中的 A、B、C、D、E、F 分别代表 6 名被评估者,汇总行的分数是各被评估者所得分数。有了表 4-14 中 5 位评估专家的评估结果,然后将各评估专家对某一被评对象评估的得分相加,就可以得到最后评估结果,表 4-14 中的评估结果是:被评对象 A、B、C、D、E、F 各自得到的总分分别是 19、9、14、13、2、18(这一分数也可以用平均分表示,即各自除以评

估专家数 5,这不改变等级评定结果),这样得到的最终排名依次是 A、F、C、D、B、E,由此就可知道:A 的管理能力和管理绩效评价等级最高,E 的管理能力和管理绩效评价等级最低。

表 4-14　采用对偶比较法制作顺序量表的过程

	评估者 1						评估者 2						评估者 3					
	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
B	A						A						A					
C	C	C					A	C					A	C				
D	A	B	C				A	D	D				D	D	C			
E	A	B	C	D			A	B	E	D			A	B	C	D		
F	F	F	F	D	F		F	F	C	D	F		A	F	F	D	F	
汇总	3	2	4	2	0	4	4	1	2	4	1	3	4	1	3	4	0	3

	评估者 4						评估者 5					
	A	B	C	D	E	F	A	B	C	D	E	F
B	A						A					
C	A	C					C	B				
D	A	B	D				A	B	C			
E	A	B	E	D			A	B	C	D		
F	F	F	C	F	F		A	F	F	F	F	
汇总	4	2	2	2	1	4	4	3	3	1	0	4

顺序量表的制作相对比较简单。当然,这类量表反映的信息也比较少,它作为一种评估量表,既无绝对零点,也无相等单位,只能测量不同事物某种排序和相对等级高低,无法测量不同事物间的精确距离,因此也无法测量它们之间的比例关系。在刚才列举的这个例子中,B 的评价等级分合计为 9,F 的评价等级分为 18,你只能说 F 明显强于 B,但你不能说 F 比 B 高出 9 个单位,也不能说 F 的等级是 B 的 2 倍。

二、等距量表与费希纳定律

等距量表没有绝对零点但有相等单位,可以确定两个刺激引起的心理量的相对大小,但不能确定二者的绝对大小和比例关系。等距量表的制订方法可以

有感觉等距法(equal sense distance method)、差别阈限法(differential threshold method)等。

感觉等距法就是根据刺激引起的感觉增加强度,找到刺激的增量序列,使得每一次增加的刺激量引起的心理量增长幅度相等。比如采用二分法比较直接、简便,操作方法是:选定两个强度不等的刺激,其引起的感觉强度有明显差异,然后调整另一个比较刺激的强度,使其引起的感觉强度正好介于两个标准刺激引起的感觉强度的中间,这样就可以把原来的感觉强度差分成相等的两个单位,接着再对两部分分别进一步二分,依此类推,就可以得到一个等距量表,该量表上的感觉量是有相等单位的。假设我们要把一个完全的黑色和完全的白色分成 8 或 16 个等距,就可以采用二分法。

也可以采用差别阈限法,即在某种强度上测量其差别阈限,然后以原来的刺激强度加上一个差别阈限的值作为标准刺激再求差别阈限;依此类推,就可以制订一个等距量表,在这个等距量表上,刺激增加一个差别阈限,感觉量就增加一个可觉差,费希纳就是采用差别阈限方法来研究刺激量与心理量的函数关系,并借此得到了有名的对数定律。

费希纳认为,差别阈限(ΔI)引起感觉量增加一个最小可觉差量(JND),因此 JND 可用作单位来测量感觉量的大小。费希纳使用差别阈限作为刺激测量单位、最小可觉差作为感觉量的测量单位,对阈上刺激进行了大量的实验研究,并对这些实验结果进行归纳与推导,得到了刺激量与心理量的对数关系定律:随着刺激强度以几何级数增加时,刺激引起的心理量则以算术级数增加,即刺激引起的心理量与刺激强度的对数成正比。对数定律的表达式为:

$$S = k \times \log R$$

式中,S 代表心理量,R 代表刺激量,k 为常数。

这一函数关系反映在二维直线坐标系中就是一个抛物形的对数关系曲线,反映在半对数坐标系中则为一条直线,如图 4-7 所示,其中的 a 线是二维直线坐标系中的曲线、b 线为半对数坐标系中的直线,二者显示的都是对数函数关系。

费希纳的研究继承了韦伯的研究成果,他的对数定律可由韦伯定律推导而来,所以费希纳定律也称为费希纳-韦伯定律(Fechner-Weber's Law)。该定律有两个假设前提:首先,假定在一种特定的感觉范畴内,差别阈限与标准刺激之

比的韦伯分数是一常数；其次，在所有刺激强度水平上，最小可觉差在心理强度上是相等的，即它是一个心理量的度量单位。

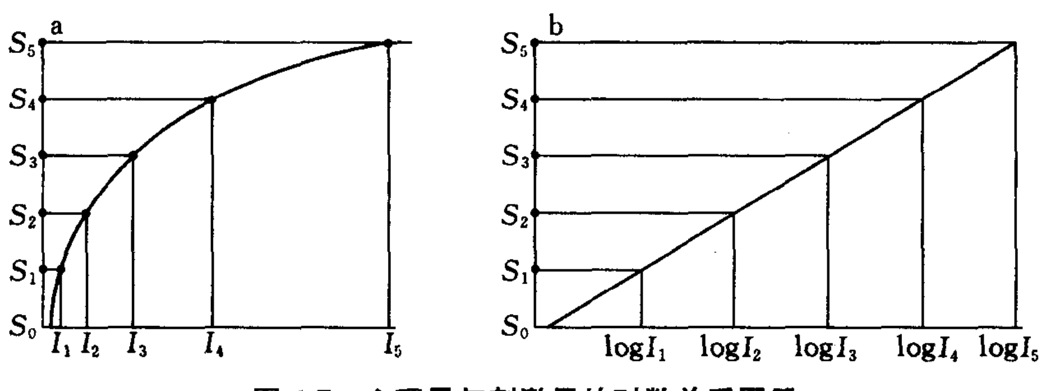


图 4-7　心理量与刺激量的对数关系图示

(a) 在二维直线坐标系中，二者的关系曲线为一抛物形对数曲线
(b) 在半对数坐标系中，二者的关系曲线为一直线

费希纳定律得到后来大量研究的证实，被看作是一种经典的心理物理学定律。但是，也有研究认为，费希纳定律只在中等刺激强度的变化范围内才能成立，在强刺激条件下就不成立了。所以，费希纳定律也得到许多批评，特别是史蒂文斯认为刺激量与心理量之间不是对数关系，而是幂函数关系，这就是有名的史蒂文斯定律。

等距量表比顺序量表的信息量大一些，它不仅能反映刺激引起的心理量的强弱顺序，而且能够使用相等单位来测量两个刺激引起的心理量的具体大小。比如，温度量表也是一个等距量表，10 摄氏度比 5 摄氏度高出 5 摄氏度，20 摄氏度比 15 摄氏度也高出 5 摄氏度，这两个温度差是一样的，但是你不能说 10 摄氏度是 5 摄氏度的 2 倍，你也不能说 0 摄氏度是没有温度(绝对零点就是绝对的“无”，0 摄氏度也是温度的一个水平，却不是“无”温度)。要想确定测量间的比例关系，还必须能够确定量表的绝对零点才行，这就是比例量表的特点和功能了。

三、比例量表与史蒂文斯定律

比例量表是既有绝对零点又有相等单位的量表，它不但可以比较各刺激引起的感觉量的相对大小，而且可以确定它们之间的比例关系。比如，我们可以设定一个响度零点，然后确定一个响度单位，就可以制订响度比例量表了。制订比例量表的直接方法是分段法。比如响度比例量表的制订是这样进行的：先规定

1 000Hz、40 分贝纯音的响度为 1 Sone，然后让被试调整一个 1 000Hz 的纯音的声级，使之听起来是 1 Sone 的响度的一半，则这时的响度就是 0.5 Sone；使之听起来是 1 Sone 的 2 倍，则这时的响度就是 2 Sone。接着再调整一个响度为 2 Sone的 2 倍，则其响度就为 4 Sone，如此不断地进行下去，就可以制订出一个响度的比例量表。

此外，还可以使用数量估计法来制作比例量表。数量估计法的操作比较简单，就是给被试呈现一系列强度不同的刺激，要求被试根据感觉强弱用具体的数字来标记每一刺激。被试可以先选择一个刺激，并根据感觉强度将其标上数字，然后将其他刺激的感觉强度与此比较，并各自标上数字以反映其强度大小。为了尽可能地使所标数字与感觉强度匹配，允许被试使用小数。

史蒂文斯(Stevens, S. S. , 1975)在使用数量估计法制作响度的比例量表中，发现了心物关系幂函数关系定律，即史蒂文斯定律(Stevens' Law)：刺激量与心理量的数量关系是幂函数关系，其表达式为：

$$S = bI^{a}$$

式中，S 代表心理量，I 代表物理量，a、b 代表常数

对该公式两端求对数，即可得到：

$$\log S = b + a\log I$$

也就是说，当刺激按几何级数增长时感觉量也按几何级数增长，比如刺激量按照 k 的倍数增加时，心理量就按照 n 的倍数增加：

$$I_i/I_0 = k \qquad S_i/S_0 = n$$

分别以 $\log I$ 和 $\log S$ 为横坐标和纵坐标作曲线得到的是一条直线，即在双对数坐标系中得到一条直线。而按照费希纳的对数关系定律，则应在半对数坐标系中得到一条直线。

一定的心理量总是由一定的物理量引起的，但是二者究竟是一种什么样的数量关系，现在并没有肯定的结论。一般认为，在心理物理学研究的历史中提出的费希纳定律和史蒂文斯定律都在一定范围内和在一定程度上揭示了心理量与物理量的关系。不过，这两个定律虽然都有实验数据的支持，但是实验者获取数据的实验方法并不相同，得到的数据性质也不同，所以他们是各自站在自己实验结果的基础上来建立物理量与心理量的关系定律的，互相否定是没有道理的。

只能说，这一关系还有待进一步研究。

阅读材料4-1

费希纳和经典心理物理学

正当通情达理、正常而年轻的赫尔姆霍兹开始积累大量证据，用以证明他的神经及心理现象的机械主义观点时，莱比锡大学一位富于幻想和神经质的中年教授也在努力工作，他要证明宇宙内所有的人、动物和植物都是由物质和灵魂构成的。居士塔夫·西奥多·费希纳(1801—1887)没有达到这个目的，不过，在收集数据来显示刺激(物质世界)和所得的感觉(意识或者灵魂的世界)之间的数学关系时，他却发展出了一套研究方法，自他以后的实验心理学家都使用这个方法来发展他的"心理物理学"。

费希纳出生于德国东南部的一个村庄，父亲是村里的牧师。这位父亲把宗教信仰与坚定不移的科学观结合了起来，就像他的儿子一样。他用上帝的语言布道，却在教堂的尖顶上安装了一根避雷针，这使村里人大惑不解，因为在当时的情况下，这份小心谨慎是对上帝信心不足的表现。上帝难道说连他自己也保护不了吗?

费希纳在莱比锡大学学医，可是1822年拿到学位后，他的兴趣转向物理学和数学。他花费了好几年的时间，把一系列用法语写的物理学及化学手册翻译成德文来赚取生活费，其翻译的书稿达到九千多页。之后，从1824年起，他开始在大学里讲授物理学，在电流方面进行了一个很大的研究项目，写了不知多少专业文章。这种忙碌的节奏使他在物理学上获得了巨大声誉，可也付出了很大代价：他开始出现头疼病，并一阵阵失去对思想的控制，这使他经常心醉神迷地忙于一大堆琐屑之事。

尽管他只有30出头而且是飞黄腾达之时——他于1833年结婚并于1834年获得全职教授职位——可是，他的身体状况却持续恶化。"我无法入睡，身心疲劳，不能思想，甚至引起我失去了对人生的信心"，他后来评述这一时期的生活。他去温泉浴场寻求治疗，可没有效果。然后，他想通过研究余象①而转移自己的

① 就是今天我们所说的视觉后象。

注意力——这是他第一次进入心理学研究。期间,他戴着太阳镜长时间观测太阳。他对余象的研究得到认可——赫尔姆霍兹,如我们所知的,使用到了他的一些数据——可是,作为观察的结果,费希纳患上了严重的恐光症,情感全线崩溃。

费希纳几乎失明,他在暗室里让自己避光,忍受着疼痛、情感压抑、无法排遣的无聊和严重的消化道毛病的折磨。(他从大学退休了,尽管只教了几年书,可还是得到了一笔养老金。)在他3年病痛生活的最低点,他把自己的房子漆成黑色,白天黑夜呆在里面,什么人也不见。各种泄药、冲浪治疗、催眠术、两种休克治疗都起不了什么作用。他还是反复不断地为琐事所烦扰;另外,他还受到两种感觉的折磨,一方面因为他感觉到自己很接近发现这个世界的秘密而喜悦,另一方面又因为得用科学方法来证明这些秘密的正确性而担心。最终,他自己慢慢好起来了,过了一阵子后,他就可看见东西而且眼睛不疼,而且还能与人讲话。他在好几个月的时间里第一次到花园去散步时,花儿看上去更明亮,色彩更鲜艳,比以前更美丽,他感觉到自己对这些东西施予了内在的光,他立即就抓住了这一点的意义:

"我毫不怀疑我已经发现了花朵的灵魂,并以我极奇怪的、受到魔力影响的情绪想到:这是躲藏在这个世界的隔板之后的花园。整个地球和它的球体本身只是这个花园周围的一道篱笆,是为了挡住仍然在外面等待着的人们。"费希纳不久写了一本书,讨论植物的精神生活,在余下的几年里,他寻找办法来宣传自己的泛灵论,即意识与物质在整个世界里是共同存在的。

正是这个神秘的信仰使费希纳进行他那具有历史意义的实验心理学研究。1850年10月22日的早晨,他躺在床上考虑如何向机械论者证明,意识和肉体是一个基本统一体的两个方面,这时,一道灵光一闪:如果他可以显示在刺激的强度与它们产生的感觉强度之间存在数学关系,则他就显示了灵肉的统一。或者在费希纳看来是这样的。这个推理的逻辑也许会逃过非神秘主义者。可是,他倒是提出了一个非常有效和重要的问题,即意识感知外部世界的准确性:刺激的强度与它产生的感觉之间是否存在前后一致的数学关系?从直觉上来说,可能是这样的:光线越强,我们看上去就越觉得亮。可是,如果你让光的亮度翻一倍,感觉的强度是否也强一倍呢?或者某种别的、好像是真实的关系存在着呢?

费希纳接受过物理学和数学的训练,他感觉到,当刺激的强度增大时,它应该要求更大的差别(绝对值上的差别)来产生一定大小的感觉增大。从数学上来说,刺激在强度上的几何增大会导致感觉的算术增大。一项临时的示意:按照传递到耳朵上的能量,一阵雷声的响声比日常谈话的声音要响好多倍;按分贝——

分贝是指人耳能够分辨的最小响度差别——来说，它只是响两倍而已。

为了通过实验确立他的直觉，费希纳得解决一个看上去无法解决的问题：他可以很容易地测量刺激强度，可是感觉是个主观的东西，而且无法测量。他推想，尽管他不能直接地观察和测量感觉，但他可以通过灵敏度的指导而间接地做到。他可以确定在感觉者刚刚能够注意到的、任何水平上最小的刺激强度增大。因为“刚刚能够注意到”在任何水平上都意味着同一个东西，这将是一个感觉的测量单位，借此可以与产生这种意识所必需的刺激的增加进行比较。

费希纳后来说，他不是从韦伯那里得到这个想法的，尽管韦伯是自己以前的老师，他在“刚好能够注意到的差别”上所做的工作几年以前就发表了。可是，他迅速意识到，他会使用到韦伯定律的推广形式。韦伯已经发现两项刚刚能够注意到的差别刺激之间的比率是一个常数，不管这种刺激的大小如何。费希纳却说，尽管两项刺激之间绝对的差别随着刺激的强度增大而增大，可感觉者对一种刚刚可以注意到的差别的感觉却仍然是一样的。

费希纳后来写道，想象一下，你用太阳镜看着天空，并在刚刚可以注意到的天空的背景上挑一片云来看。然后，你戴上一副黑得多的眼镜，云彩没有消失，但几乎只能够注意到：因为，尽管强度的绝对水平在透过黑镜片的过滤后会低得多，但云彩与天空之间的强度比却没有变。

为了表达刺激强度与感觉强度之间的关系，费希纳从数学上转变了韦伯定律，重新加以调整然后写出来：$S = k\log R$。

这意思是说，分段式的感觉强度增大是刺激强度翻倍的结果(乘以某个比率或者系数)。费希纳拼命要把这份荣誉还给他以前的老师，因此，他把这个公式称作韦伯定律——是他本人给韦伯的公式和他自己的公式命名的——可是，后来的心理学家按照这些公式各自的归属把修改后的公式叫做费希纳定律。

费希纳余下的9年花在辛苦的实验工作当中，收集着大量能够确证这个定律的数据。尽管他的性格当中有一些神秘主义者和诗人的气息，可在实验中，他是一位有强制力和严厉的研究者的榜样。他不知疲劳地让受试者们举起重物，注视光源，听各种杂音和音调，观察彩色样本等等，并宣布它们是不同的或是同样的。在这些年里，他对每种刺激的强度进行了范围广泛的实验，使用到了测量这些判断的三种方法。仅在这些方法中的一种里面，他便列出了24 576种判断的表格和计算结果。他认为，第一次系统地探索物理和心理学王国之间的数量关系，这是一种新的科学专业，因此命名为“心理物理学”。

在他使用过的三种实验测量方法当中，他从前人那里借来了两种并使之完善，然后自己发明了第三种。直到当时为止，没有人曾使用过这种极仔细、可能准确控制的数量测量方法来探索心理学的反应。他的方法很快就被广泛接受，而且在今天心理学的每一个实验室里还在经常使用着。

一种是极限的方法，费希纳本人管它叫“仅仅可以注意到的差别法”。为了确定一个刺激的临界值，实验者一次提供一个刺激，从最低的量开始逐渐加大强度，直到受试者可以感受到刺激为止；为了确定仅仅可以注意到的差别，实验者提供一个“标准的”刺激和一个“比较用的”刺激，以小幅度增大两种之间的差别，直到受试者说可以感觉到为止。

第二个方法是常量刺激法，费希纳本人叫它“正确和错误情况法”。实验者一次又一次提供等量的刺激——在临界值上的单个刺激或者成对相似的刺激。受试者回答说“有了”(意思是说，他感受到了它，或者二者有了不同)，或者说“没有”(表示他没有感觉到，或者两个刺激没有什么不同)，受试者的回答得出平均值，而这些平均值会指明，在任何指定的刺激水平，或者两个刺激之间的差别上，受试者有没有可能会感觉到这些刺激，或者两个刺激之间的差别。

第三个方法，即费希纳本人的创造性贡献，叫做调节法，他把它叫做“平均错误法”。实验者或者受试者调节比较刺激，直到它好像(对受试者来说)与标准刺激相等。这边或者那边总是会有一些错误存在的，不管多么微小。每一次错误都记录下来，许多次试验过后，把平均错误加起来，它也是仅仅可以注意到的差别的尺度。这个方法确立了这个有用的原则，即测量数据可变性可以跟测量集中趋势一样得出信息来。

1860年，费希纳出版了两卷本的《心理物理学纲要》，把他的研究成果公之于世。他已经59岁了，在这个年纪，科学家一般很少会拿出有创见的东西来；可是，《心理物理学纲要》的确是富于创见的，并立即产生了很大的影响。兴趣是浓厚而且广泛的——不是对他信奉的泛灵论，而是对它的实验和计量方法学。如波林论及费希纳的成败时曾说过，“他攻击物质主义的铜墙铁壁，但又因测量了感觉而受到赞美”。确切地说，有些心理学家认为，心理物理学的方法学是一个可怕的话题。许多年以后，伟大的威廉·詹姆斯写道：

“如果像他这样一位可敬的老人会用他的怪想使这门科学永久地背上负担，而且，在一个充满各种容易产生成果的、引人注目的事物的世界里，如果迫使未来的学者们在这些繁杂的田地里耕耘，不仅要去研究他的作品，而且得研究那些

反对他的更枯燥的作品，那可真是一件要命的事。”

可是，很多人并不持这样的看法。尽管对费希纳说仅仅可以注意到的差别都是相等的这个假想的有效性存在激烈的争论，可是，他的方法一般却都认为是天才的突破。对刺激和反应两者之间的关系进行计量研究的时机已经成熟；许多心理学家几乎立即就开始利用费希纳的三个方法了，这三个方法将肉体的生理机构与它们所引起的主观的经验连接起来了。（费希纳本人尽管仍然还在写文章为他的心理物理学辩护，可是，他把漫长余生的大部分都贡献给美学、轻度异常现象、统计学和泛心灵哲学的研究上了。）

后来的心理学家们已经发现有错误，甚至驳倒了他的发现中的每个地方，可是，他的方法不仅仍然有用，而且是感觉测量中最基本的方法。波林总结了费希纳相互矛盾的成就：

“没有费希纳……也许仍然会有一种实验心理学……可是，在实验体中，却不可能出现如此广泛的科学范畴，因为，如果测量不能成为科学的工具之一，则我们很难认为某个课题是符合科学的。因为他所做的事情和他做这些事情的时代，费希纳创立了实验计量心理学，并把这门学问从其原来的途径搬回来导入了正轨。人们也许可以称他为实验心理学‘之父’，或者，人们也许会把这个称号送给冯特。这没有什么关系。费希纳种下了肥沃的思想之种，它生长起来，并带来了丰硕的成果。”

——资料来源：《心理学的故事》（李斯，1999）①

建议阅读文献

1. 杨治良主编：《基础实验心理学》，甘肃人民出版社 1988 年版，第 90—148 页。
2. 杨治良编著：《实验心理学》，浙江教育出版社 1998 年版。
3. 孟庆茂、常建华编著：《实验心理学》，北京师范大学出版社 2001 年版。

复习思考题

1. 如何理解感觉阈限及其操作定义、最小变化法、恒定刺激法、平均差误法、

① ［美］墨顿·亨特著，李斯译：《心理学的故事》，海南出版社 1999 年版。

25%的差别阈限(或75%的差别阈限)、信号检测论、似然比、辨别力指数、信号接收者操作特征曲线(ROC曲线)、有无法实验、评价法实验、心理量表、顺序量表、等距量表、比例量表、对数定律、幂函数定律?

2. 如何理解感觉阈限的本质?
3. 如何使用最小变化法测定绝对感觉阈限和差别感觉阈限?
4. 如何利用恒定刺激法测定绝对感觉阈限和差别感觉阈限?
5. 如何利用平均差误法测定绝对感觉阈限和差别感觉阈限?
6. 如何进行有无法和评价法的实验,二者有什么区别和联系?
7. 如何根据信号检测实验的数据计算似然比和辨别力指数?
8. 信号检测实验中,被试信号判断的标准如何测量?其影响因素有哪些?
9. 三种经典心理物理学方法各有什么优缺点?
10. 与经典心理物理学相比,使用信号检测论测量感受性的优点是什么?

第五章
反应时间

本章内容提要

反应时间是心理学研究中一个重要而特殊的领域，一方面它作为心理学基础研究和应用研究的重要方面，另一方面它作为最常用的实验因变量，成为许多研究范式的构成成分。本章首先介绍反应时间研究的简史，并借此讨论反应时间的基本概念、主要种类、测试方法。其次，广泛讨论反应时间的影响因素，这主要包括两大方面：主观因素如年龄、性别、人性差异、高级神经活动类型、动机、身心状态等，客观因素如刺激的性质、强度、持续时间、呈现模式等等。反应时间测量在心理学研究和实践领域中都具有广泛应用，对于心理学研究来说，减法反应时法、加因素法、开窗实验是三种典型研究范式；对于实践应用来说，在人才选拔、工业设计和交通运输领域，反应时间测量都具有直接和间接的意义。

反应时间(reaction time,简称 RT)就是个体从接受刺激作用到作出相应反应的时间间距，也叫反应潜伏期。反应时间在实验心理学中的地位很特殊，它本身不是心理现象，但却是实验心理学最早的实验研究课题之一，也是心理学实验研究中迄今最为常用的反应变量，成为追索人类信息加工内部机制的一根有效“探针”。本章中，我们不仅要回顾实验心理学初建前后的反应时间研究，而且还要考察反应时间在信息加工心理学研究中的特殊价值，以及在多个不同实践领域中的应用。

第一节　反应时间研究简史

心理学将反应时间作为专题进行系统的实验研究开始于 1879 年。在莱比

锡大学一所叫孔维特的破旧公寓楼里，一天，冯特教授与他的两名学生，弗里德里奇和霍尔，在这所小楼的一间小房子里摆弄着一套简陋器具，他们要用这套器具为弗里德里奇收集博士论文的数据，他的博士论文的标题就叫《知觉的长度》①——实际上，他们测量的就是听觉刺激条件下的简单反应时间。这大概是心理学家中最早的反应时间实验研究。何谓简单反应时间呢？我们暂且将此问题搁置一边，去追溯反应时间研究的早期历史，这历史是从天文学和生理学那里开始的。

一、反应时间研究的早期历史

1796 年，格林威治天文台的台长马斯基林（Maskelyne）发现助手金内布鲁克（Kinebrook）观测的一些星体经过子午线的时间总是比自己慢 1 秒钟左右，他因此认为这名助手的工作很不严谨，非常恼火地辞退了他。二十几年后，德国天文学家贝塞尔（Bessel）注意到这一事件的档案资料，洞察了这一事件的特殊意义，开始对其进行思考和系统研究。他对当时天文学家采用"眼耳法"（eye-ear method）观察星体运行过程的数据进行比较，也看到了不同观测者记录时间的明显差异。"眼耳法"观测星体运行过程的方法是：观测者在大致知道某一星体经过望远镜中法线时间的情况下，就可以等到星体接近中法线时，先看钟表指针所指的时间位置，然后边通过望远镜观察星体的运动，边用耳听着钟表的"咔哒"声并计数，这样就可以记录星体经过望远镜中法线的时间。1823 年，贝塞尔发表了他的研究发现，对自己与另一位天文学家阿格兰德（Angalander）观察 7 颗星体经过望远镜中法线的时间进行了比较，两人每一次的观测时间几乎都相差1.2 秒左右，是一个相对稳定的值，这就是著名的贝塞尔-阿格兰德人差方程式②，即：$B-A=1.233$ 秒。

这一发现很快引起其他天文学家的兴趣，纷纷对之进行研究。不过天文学的研究更多地从时间测量本身的角度和时间测量工具的设计与制造的角度开展，还较少具有"心理学"的性质。生理学家随后开展的"生理时间"的研究则带有明显的"心理学"性质，特别是荷兰生理学家唐德斯（Franciscus Cornelis Don-

① ［美］墨顿·亨特著，李斯译：《心理学的故事》，海南出版社 1999 年版。
② 张庆云：《基础实验心理学》，河南大学出版社 1993 年版，第 130 页。

ders, 1818—1889)的反应时间测定方法,简直就是现代信息加工心理学方法的雏形。

首先,在 1850 年,德国生理学家赫尔姆霍兹(Hermann von Helmholtz, 1821—1894)成功地测出蛙运动神经的神经传导速度约为 26 米/秒。此后,他又在人体不同部位的皮肤接上电极,施加电击,并要求被试在感受到电击时立即用手作按键反应,由此推断出人的神经传导速度约为 60 米/秒。这一结果后来虽被证明不是非常准确,但这一测量本身在心理科学诞生中具有重要的奠基性价值,它使人们相信神经过程和心理活动的时间特性是可以测量的。

根据测定,赫尔姆霍兹认为神经传导占用的时间是很有限的。但是,完成整个反应过程需要的时间却比较长,那么这些时间被耗费在什么环节呢? 引出此类课题是很自然的事。1868 年,荷兰生理学家唐德斯第一次试图把心理因素加进去以研究其对反应时间的影响。简单反应时间、选择反应时间、辨别反应时间就是唐德斯提出来的一些实验范式。通过研究,唐德斯认为,一个复杂反应只是在一个简单反应上加一些别的动作,这些动作所用的时间可用反应的全部时间减去简单反应需要的时间求得,该时间差就是加进去的这种心理活动需要的时间。与此同时,奥地利生理学家埃克斯纳(Exner, 1873)指出被试在反应时间实验中的定势(或心向)(set)的重要性,并最先使用"反应时间"的概念。

1868 年,唐德斯在《关于心理过程的速度》一文中发表了他的反应时间研究结果。他首先测定了三种不同的反应时间:

A-反应时间:只有一个反应对应于一个特定的刺激,当该刺激呈现时,被试就立即作出规定的反应,这时测得的反应时间就是简单反应时间(simple reaction time)。

B-反应时间:有多个反应,各自对应于一个指定的刺激。当某一刺激呈现时,被试必须作出与之对应的反应,这时测得的反应时间就叫做选择反应时(choice reaction time)。选择反应时间包含简单反应时间、辨别刺激的时间和选择反应的时间。

C-反应时间:只有一个反应,它仅与多个刺激中的某一刺激相对应。只有当对应的这个刺激出现时,被试才作出规定的反应,这时测得的反应时间叫做辨别反应时间(identification reaction time),其中包含简单反应时间和辨别刺激的时间。

有了上述三种反应时间的数据，就可以从中计算出辨别刺激过程的时间和选择反应过程的时间。辨别刺激的时间可以用C-反应时间减去A-反应时间得到，选择反应的时间则可以用B-反应时间减去C-反应时间得到。唐德斯的研究，是从已知的心理过程测量其所需要的时间，后来这一研究范式扩展到用反应时间差来推断信息加工的过程，成为现代信息加工心理学研究的典型范式。

把反应时间作为心理学研究课题者首推冯特，因为也是从冯特开始，心理学才被看作一门独立科学。冯特在建立实验心理学的初期，就注意到唐德斯的工作，认为唐德斯指出了实验心理学的一条重要途径，即心理活动的时间测定工作。冯特的学生对简单与复杂的反应时间进行了一系列的研究，但他们在注意、知觉、联想和选择等过程上却未测出确切的反应时间。在冯特的早期学生中后来有的建立了反应时间研究的专门实验室，其中卡特尔(James MeKeen Cattell, 1860—1944)开展了许多反应时间研究实验，他认为被试在做简单反应测验时，其注意力完全集中于那个将出现的刺激和那个将要做出动作的手指。当刺激到来时，眼睛——大脑——手指之间的神经通路早已准备好了，反应时间就短。在辨别和选择反应的实验中，需要有更多的神经通路接通的准备，这时被试的心理状态就比较复杂，会产生焦虑、怀疑等复杂的心理状态，所以反应时间就会延长。从卡特尔开始以后的研究已不在于分析它的原因，而是转向测量技术的改进，以及深入到实际的运用领域中去。

心理学家在总结反应时间研究的早期历史时，把自1850年赫尔姆霍兹的研究至1969斯腾伯格(Sternberg, S.)的研究之间长达一百多年的时间称为唐德斯反应时间ABC时期。这是反应时间研究的第一阶段，这一阶段的方法学的核心是减法反应时间法。1969年以后，由于认知心理学研究中普遍将反应时间作为因变量，使反应时间的研究进入新阶段，被称作加法反应时间时期①。

二、认知研究中的反应时间实验

正如前文所述，反应时间可以间接反映信息加工的内部过程，所以它成为认知心理学研究中的基本变量，甚至可以说其一度成为认知心理学的方法学基础。认知心理学中的反应时间法实验的发展经历了三个阶段：减法反应时间法、加因

---

① 孟庆茂、常建华：《实验心理学》，北京师范大学出版社1999年版，第140页。

素法、开窗实验。下面,我们通过典型实验来对之进行介绍。

(一) 减法反应时间法

减法反应时间法(method of minus reaction time),也称为减法法则(minus rule),是由唐德斯的实验范式扩展而来的,后来成为信息加工心理学的基本研究方法。有许多关于信息加工内部机制的发现就是通过减法反应时间完成的,比如心理旋转、短时记忆编码方式、语义层次网络模型等。减法反应时间法的基本原理是:当两个信息加工系列具有包含和被包含关系时,即其中一个信息加工系列除含有另一个信息加工系列的所有过程以外,还存在一个独特的信息加工阶段或过程,这两个加工系列需要的时间差就是这个独特的信息加工阶段或过程需要的时间。如辨别反应包含简单反应的全部加工阶段,同时它还有一个刺激分辨的心理加工阶段是简单反应所没有的,那么通过反应时间的相减就可以得到刺激分辨的心理加工需要的时间。很明显,心理加工越复杂,需要的加工时间就越长。

为理解这一方法的实质和实验操作过程,现介绍两个经典实验:心理旋转实验和短时记忆编码之争实验。

1. 心理旋转实验

所谓心理旋转(mental rotation),是指在空间知觉加工过程中进行的一种心理上的旋转操作,从而获得正确知觉经验的历程。那么,这种心理旋转是否真的存在呢? 20 世纪 70 年代初,库珀和谢泼德[①]用减法反应时间法为心理旋转提供了有力的支持证据。库珀和谢泼德使用左右不对称的字母或数字(如,R、J、G、2、5、7 等)以及这些字母或数字的镜像作为实验材料,以六种不同的倾斜角度呈现,就构成了 12 种不同的刺激材料,如图 5-1 所示。

实验中在电脑显示器上按随机顺序给被试呈现图 5-1 中的 12 种刺激材料,每次呈现时要求被试判断其是正写的字母或数字,还是反写的字母或数字,并要求其尽快地按相应的键作出反应,电脑自动记录其反应时间和正误。多次重复之后,计算被试在每一种倾斜角刺激条件下的平均反应时间(将同一种倾斜角的正写字母或数字、反写字母或数字的反应时间相加平均),得到了刺激倾斜角度与被试反应时间之间的关系曲线,如图 5-2 所示。

---

① Cooper, L. A. & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase(Ed.), *Visual information processing*. New York: Academic Press, 75—176.

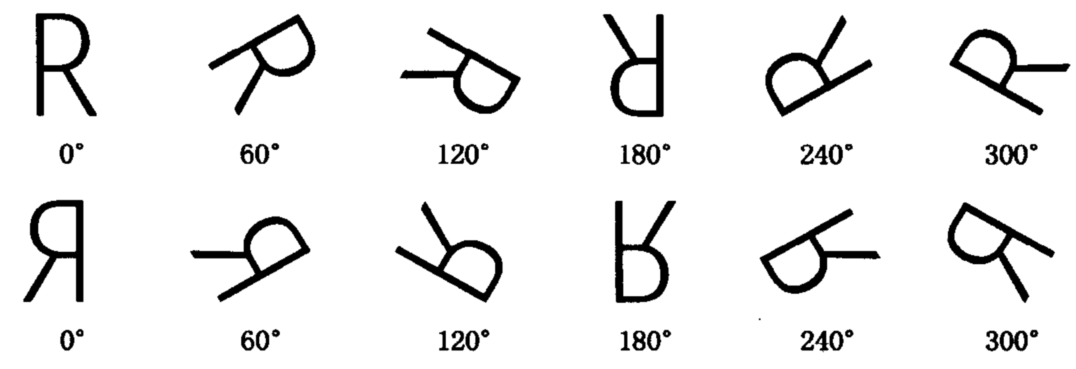


图 5-1 库珀和谢泼德(Cooper & Shepard, 1973)使用的实验材料示例

上行是六种倾斜角度的大写字母 R(也称为正 R)

下行是六种倾斜角度的字母 R 的镜像(也称为反 R)

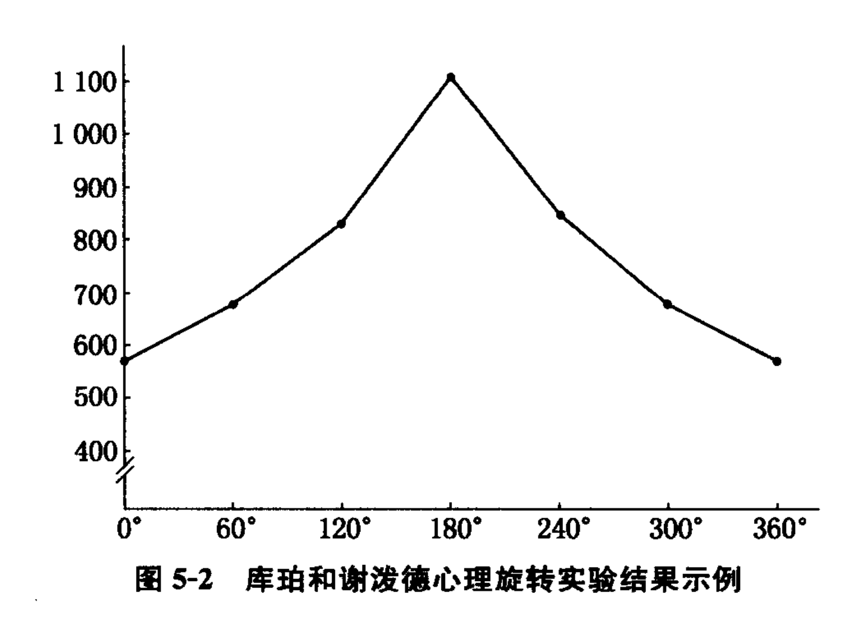


图 5-2 库珀和谢泼德心理旋转实验结果示例

从图 5-2 看出,当呈现材料的倾斜角度不同时,被试的反应时间也不同,而且正立位置(正 R 和反 R 的倾斜角为 0 度)的反应时间最短、倒立位置(正 R 和反 R 的倾斜角为 180 度)的反应时间最长,整个变化的关系曲线呈现出以 180 度为中心的对称性。这一结果如何解释呢? 库珀等认为,刺激材料正立时,被试可以快速地将刺激材料的知觉印象与相应的记忆表象进行对照,对刺激获得快速识别,其反应时间就短;当刺激材料以一定的倾斜角度呈现时,其获得的知觉印象不能立即与相应的记忆表象匹配,所以需要将直接获得的知觉印象进行心理操作,即对其进行心理旋转,等其旋转到正立位置时就可以和相应的记忆表象匹配,获得识别,所以这时的反应时间加长了。自然,倾斜角度越大,心理旋转的角度也越大,消耗的时间也越多,反应时间也就越多。但是,为什么倾斜角度超过 180 度时,角度增加反应时间反倒缩短呢? 因为根据心理加工的经济性原则,当

刺激材料倾斜角超过 180 度时,其较短的旋转过程应该是顺时针的。具体地说,当刺激倾斜角为 120 度时,其按逆时针旋转到正立位置是最近的;当刺激倾斜角为 240 度时,其按顺时针旋转到正立位置是最近的,这两种情况下心理旋转的角度应该相等,反应时间也应该是相等的,这种分析得到了实验结果的支持。

概括来说,库珀和谢泼德的实验证实了心理旋转假说。当以一定倾斜角呈现刺激时,被试获得最初的知觉印象后,对该知觉印象进行心理的旋转操作,获得与相应记忆表象匹配的知觉经验后方可完成识别和判断,心理旋转过程消耗的时间等于该种条件下的反应时间减去正立位置刺激条件下的反应时间。

2. **短时记忆编码之争**

在 20 世纪 60 年代以前,人们一般认为短时记忆信息以听觉编码形式储存。但是 70 年代初波斯纳等人①的实验表明,短时记忆的信息存在视觉编码,其根据就是减法反应时间法实验的结果。实验的具体做法是:给被试并排呈现两个字母,这两个字母可以同时给被试看,也可以间隔一个短的时间间距先后给被试看,要求被试指出这一对字母是否相同并按键作出反应,记录反应时间。所用的字母包括:读音和书写都相同的字母(如 A-A)、读音同书写不同的字母(如 A-a)、读音不同的字母(如 A-B 或 A-b),读音相同的都应判断为同一字母,被试应作出"是"的判断;读音不同的字母就是不相同的字母,被试应作出"否"的判断。在有时间间隔的实验中,时间间隔包括 0.5 秒和 1 秒、1 秒和 2 秒,而且被试的反应时间是指从第二个字母呈现到其作出按键反应之间的时间间距。波斯纳等人的实验结果如图 5-3 所示(图中只显示了 A-A 和 A-a 刺激系列的实验结果,A-B 和 A-b 等刺激系列的实验结果未列出)。图 5-3 所示的实验结果显示,当两个字母同时呈现时,A-A 刺激条件下的反应比 A-a 刺激条件下的反应要快得多;当两个字母先后呈现时,A-A 刺激条件下的反应时间开始增加,但 A-a 刺激条件下的反应时间变化不大,而且两个字母时间间隔越大,A-A 刺激条件下的反应时间增加更多,比如在间隔 1 秒以上时,就与 A-a 刺激条件下的反应时间接近。波斯纳等人认为,视觉刺激的短时记忆编码既有听觉编码,也有视觉编码,而且先有视觉编码,然后向听觉编码转换。在 A-a 刺激条件下,两个字母的听觉编码相同

① Posner, M. I., Boies, S. J., Eichelman, W. H. & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 79(1, Pt. 2), 1—16.

但视觉编码不同，所以在接受到刺激后，前期形成的视觉编码并不匹配，所以还不能作出是否为同一字母的决策，于是编码形式开始向听觉编码转换，转换基本完成时就可在听觉编码上实现匹配，被试作"是"的回答，这一过程消耗的时间比较长，所以反应时间就比较大。如果 A-a 采用间隔呈现，其决策也是在听觉编码匹配后完成的，所以反应时间与同时呈现时的反应时间相比差异不大。对于 A-A 刺激系列，情况则有所不同：当同时呈现时，两个字母首先就会在视觉编码上完全匹配，被试在刺激编码还尚未向听觉编码转换时就可决策；如果两个字母先后呈现，当第一个字母呈现后获得视觉编码时，第二个字母尚未出现，这时第一个字母的视觉编码开始向听觉编码转换，第二个字母随后出现时形成的视觉编码就与短时记忆中前一个字母的编码形式无法匹配（前一个字母在一定程度上已经是听觉编码形式了），所以第二个字母也要向听觉编码转换才能与之匹配，需要的时间就增加，而且字母呈现的间隔越大，第一个字母向听觉编码转换越彻底，反应时间就越长。当两个刺激间隔时间达到 1 秒以上时，A-A 系列也基本上都是在听觉编码上进行匹配的，所以其反应时间与 A-a 系列相近。

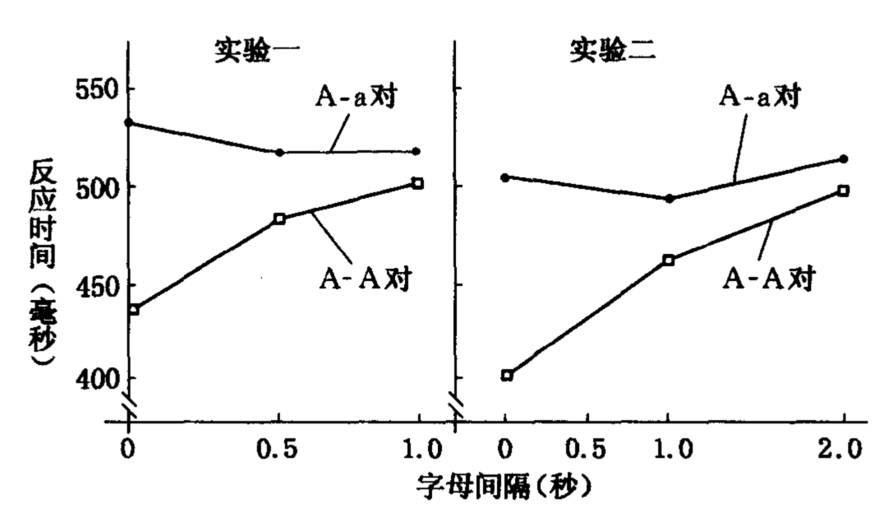


图 5-3　反应时间是字母时间间隔的函数(Posner et al.，1969)

波斯纳等人的反应时间实验证实了某些短时记忆信息可以有视觉编码和听觉编码两个连续的阶段，这也是信息加工心理学中的重大发现。

以上两个实验都是减法反应时间法的典型范例。减法反应时间法实验的逻辑是：安排两种反应作业，其中一个作业包含另一个作业所没有的一个加工阶段，其他方面均相同，从这两个反应时间之差来判定此加工阶段。这种实验在原则上是合理的，在实践上是可行的。信息加工心理学也主要用减法反应时间实

验提供的数据来推论人脑内部的信息加工过程。

但是这种实验方法也有弱点，它要求实验者对实验任务引起的刺激与反应之间的一系列心理过程有较为明确的认识，并且要求两个相减的任务中共有的心理过程要严格匹配，这一般是很难的。此外，减法反应时间法还有一个最致命的缺陷，即它的假设前提认为信息加工是系列化的，也就是由一系列前后相连的多个加工阶段构成的，这一观点受到越来越多研究发现的挑战。这些弱点大大限制了减法反应时间法的广泛使用。

(二) 加因素法

20 世纪中期，斯腾伯格(Sternberg, 1969)发展了唐德斯的减法反应时间法，提出了加法法则，被称为加因素法(additive factor method)[①]。这种方法并不是对减法反应时间法的否定，而是减法反应时间法的发展和延伸。加法反应时间实验认为完成一个作业需要的时间是一系列信息加工阶段分别需要的时间的总和，如果发现可以影响完成作业需要时间的一些因素，那么单独地或成对地应用这些因素进行实验，就可以观察到完成作业时间的变化。加因素法实验的基本逻辑假设是：如果两个因素对某一信息加工过程的影响具有交互性(交互效应明显)，它们导致的信息加工过程的时间变化不具有可加性，那么这两个因素影响的是信息加工过程中的同一个阶段；如果两个因素对某一信息加工过程的影响是相互独立的(交互效应不显著)，它们导致的信息加工过程的时间变化具有可加性，那么这两个因素影响的是这一信息加工过程中相互独立的两个不同阶段。按照这一逻辑，通过单变量和多变量实验，测量被试完成作业的反应时间的变化，就可以探测到影响某一信息加工过程的独立变量，进而可以推断这一信息加工的阶段。在加因素实验中，重要的不是测量每个阶段的加工时间，而是要探测加工的阶段及顺序。加因素法假设，当两个因素影响两个不同的加工阶段时，它们将对总反应时间产生独立的效应，即不管一个因素的水平如何变化，另一个因素对反应时间的影响都是恒定的，这就是因素影响效应的相加性。

1969 年，斯腾伯格使用加因素法进行了一个有名的短时记忆信息提取方式的经典实验研究。斯腾伯格先给被试看 1～6 个数字(识记项目)，然后看一个数字(测试项目)并同时开始计时，要求被试判定此测试数字是不是刚才识记过的，

① Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta Psychologica*, 31:276—315.

通过按键作出"是"或"否"反应,记录这期间的时间即反应时间。斯腾伯格按照加因素法,改变实验过程中的刺激因素,得到一系列不同条件下的反应时间。根据这些反应时间的变化关系,斯腾伯格探测出被试对识记数字系列的短时记忆信息进行提取的阶段和独立影响因素:四个独立的加工阶段为刺激编码阶段、顺序编码阶段、二择一的决策阶段和反应组织阶段,对应的独立的影响因素分别为测试项目的质量、识记项目的数量(记忆集大小)、反应类型(即肯定和否定)、反应类型的相对频率。这一研究结论可表示成图 5-4 的形式。

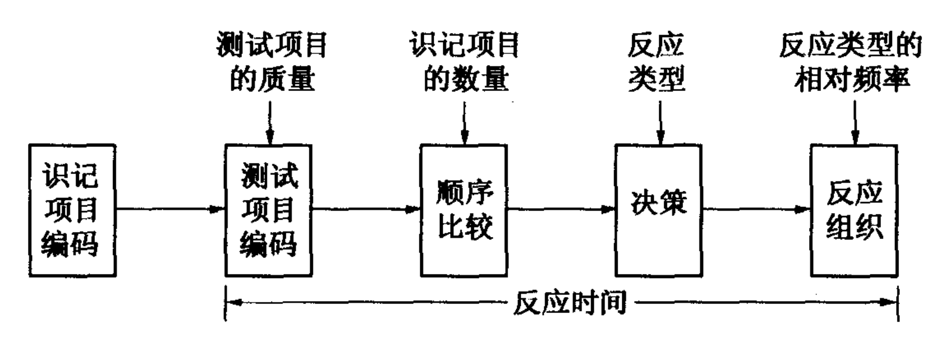


图 5-4　斯腾伯格(Sternberg, 1969)的加因素实验:短时记忆信息提取的过程及影响因素

实验中得到的结果主要包括:(1)测试项目的质量较低时(残缺的或模糊的),反应时间较长;(2)记忆集越大时,反应时间越长,表明顺序比较阶段主要是系列扫描的;(3)"是"反应快于"否"反应,这说明两类反应的难度不同,"否"反应难度较大;(4)反应类型的相对概率影响反应时间,一般某种反应概率增加时,其对应的反应时间会缩短。

斯腾伯格的加因素法引起了许多心理学家的兴趣,激发了一系列类似的研究,同时也招致许多质疑和批评。一些心理学家指出了加因素法实验中的三个致命之处:首先,加因素法实验是以系列加工假设为前提的,这对其应用范围是一个制约,因为认知心理学的研究提供了越来越多的证据证实信息加工过程并非只是系列性的,而是存在平行加工(parallel processing)。第二,在加因素法实验中,变量操纵带来的反应时间变化的可加性或不可加性,能否作为确认信息加工阶段的依据,似乎还缺乏足够的论证,这是加法反应时间法的基本逻辑。既然对其基本逻辑还存在怀疑,那么加因素法本身也当然会被怀疑。第三,加因素法本身还没有解决如何确定加工阶段的顺序问题。所有这些批评意见是有道理的,值得重视。但不管怎么说,加因素法是对反应时间法的一次发展,也因此促进了认知心理学的发展。

（三）开窗实验

开窗(open window)实验是心理学中发展较晚的一种方法，这种方法既与减法反应时间法有相似之处，也与加法反应时间法有共同的地方，但又有其独特性。减法反应时间法和加因素法都很难直接得到某个特定加工阶段需要的时间，都需要通过间接比较才能得到，并且这个加工阶段还要通过严密推理才能确认。如果能直接测量每个加工阶段的时间，而且也能明显地看出这些加工阶段，就好像打开窗户一样，故称其为开窗实验。开窗实验的典型例证是汉密尔顿(Hamilton, 1977)和霍克(Hockey, 1981)开展的字母转换实验。

研究者给被试呈现1～4个字母，并在字母后面加上一个数字，如“F＋3”、“GNEC＋4”、“FQ＋2”等。当给被试呈现“FQ＋2”时，就是要求被试转换并报告出与F和Q对应的后边第二个字母，即“HS”。实验中要求被试要进行出声转换，而且是逐个字母进行转换，等到要求转换的字母全部转换完成后将全部结果一同报告出来。这一实验需要在电脑程序控制下完成，我们以“GNEC＋4”转换任务为例，并结合图5-5来说明字母转换实验的过程和结果。①

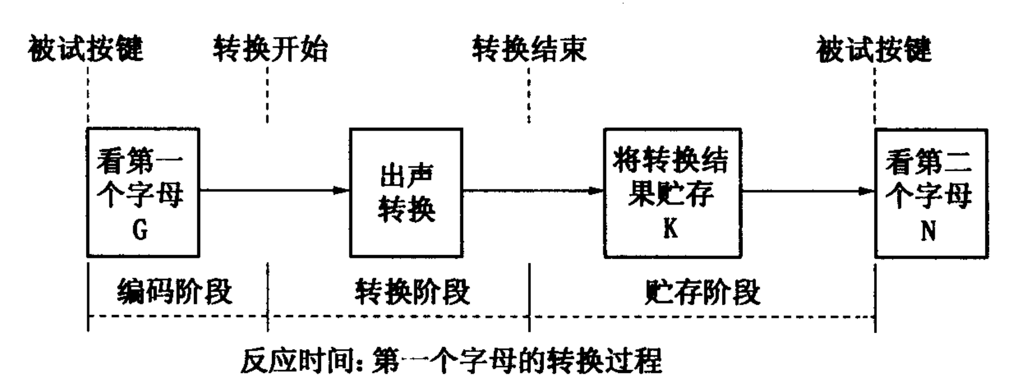


图5-5　开窗实验的过程和结果(以GNEC＋4为例)

如图5-5所示，被试做好准备后按键(电脑在程序控制下开始计时)，在显示器上呈现出第一个字母，被试就开始识别这一字母，并在字母表上寻找该字母的位置，这是对刺激的编码过程；找到其在字母表上的位置后，就立即开始出声转换，转换出相应的字母就是转换结束的时间，然后就进入贮存过程；被试觉得记住了转换结果后就再次按键看下一个字母，如此循环进行，直到四个字母均转换完成后报告出所有转换结果。在实验过程中，进行时间记录，并标记出每一次出声转换的开始时间和结束时间。这一实验过程，可以直接记录字母转换的过程，

① 转引自：张庆云：《基础实验心理学》，河南大学出版社1993年版，第152—154页。

其经历的信息加工阶段、每一阶段消耗的时间都可以直接测得。具体地说，字母转换任务的信息加工包括三个阶段：从按键呈现字母到开始出声转换是刺激编码阶段；从出声转换开始到转换结束，这是信息转换阶段；从出声转换结束到再一次按键是转换结果的贮存阶段。

当然，事实并非如此简单，每个加工阶段并非如实验过程直接看到的那样界限分明，一个加工阶段的出现有可能是对前一阶段的复查，而贮存阶段也会包含对以前转换结果的提取、复述和归并。所以，只能把开窗实验看作是对信息加工过程的简化的、直接的测量。

认知心理学研究中的反应时间实验法将反应时间置于很重要的地位，也因此促进了反应时间相关因素的研究和反应时间测量技术的发展。但同时也必须看到，上述反应时间法实验均有值得进一步探讨的问题，比如反应时间实验均把信息加工看成是由多个界限分明、相互独立、系列排列的加工阶段组成的，这种观点越来越受到挑战。

第二节　反应时间测量及影响因素

通过反应时间探测心理加工的过程与机制成为心理学研究中的一种有用范式，反应时间测量也因此成为心理学的一种基本实验技术。

一、反应时间的测量

计时活动和计时工具的发明与制造均可追溯到很久远的年代，但是反应时间具有特殊的意义，它特指人在刺激作用下到作出某种反应的时间间距，所以反应时间的测量装置必须包括三个基本的部分：刺激呈现装置、反应操作装置和计时装置。

（一）刺激键与反应键

刺激键(stimulus key)和反应键(response key)是控制计时器开动和停止的装置，常常使用电键来替代。不过电键存在一些缺点，往往有接触不良、噪声大的缺点，很难保证实验质量。例如，使用普通电键来完成视觉刺激条件下的反应时间，如果被试和主试都在同一个实验室内，在主试操作刺激键时发出的声音就成为一种额外刺激，被试就未必是按照视觉刺激来反应。这时就要使用不会引

起噪声的开关。在某些运动反应时间测定中,也可以采用光电开关,这种开关不用手操作,而是用被试肢体的某部分反应遮挡照射到光电管上的光线,或从遮挡处移开让光线照射到光电管上引起光电流的变化,控制计时电路。在测定言语反应时间时,还可以使用声音开关,即用声音控制计时的停止等。

随着计算机软硬件的发展,许多反应时间的测量任务可以由计算机及其程序来控制,即以显示器或耳机作为提供刺激信息的"刺激键",而以计算机键盘或鼠标作为被试操作的反应键,也可以使用声音控制来完成语音反应的反应时间测试。①

(二)计时装置

与刺激键和反应键相连接的是计时装置。心理学实验室中使用比较多的是数字计时器,它与刺激键和反应键连接,利用记录电路中的电流脉冲数来记录时间。刺激出现时,通过刺激键发出的信号激活计时器开始计时;随后被试操作反应键作出反应,通过信号终止向计数器发送脉冲信号,计时器就停止计时,这样就可以记录下刺激呈现到被试作出反应的时间间距。不过,随着实验精度或实际应用中对时间测量要求越来越高,特殊摄影和计算机自动计时被应用于反应时间的记录与测量。

特殊摄影就是使用摄像机将被试从接受刺激到作出反应的过程录制下来,必要时还可以改变拍摄的采样率和摄像画面的播放速度,非常准确地计算出某一反应的时间。而且在摄影时,也可以同时将被试的操作及计时器上的数字同时摄入画面,对时间的读取就非常方便了。对于一些特殊计时领域来说,使用摄影手段是非常必要的,比如在短跑比赛中,就可以使用摄像技术准确地记录运动员从发令枪响到起跑之间的反应时间。

(三)反应时间测量中应注意的问题

为了准确地测量反应时间,需要注意以下三个方面的问题:

第一,避免过早反应及其他错误反应

实验前,或许有些被试会对反应时间的长短不以为然,但是在实验开始后几乎所有的被试都希望自己测得的反应时间越短越好,因此实验中常常控制不住自己的手,甚至在刺激出现之前就作出了反应,或者在刺激刚刚出现,手就几乎同时将反应键按了下去,这就是所谓的"假反应"或"抢码"。如果实验中刺激呈

① 南京师范大学心理学科"E-Prime 实验室"可以非常容易地将需要记录反应时间的实验过程程序化,其刺激呈现的时间和反应时间的记录都可以在毫秒单位上实现。

现的时间间隔固定，这种“抢码”现象就更为频繁。为了避免这种现象的发生，通常可采用两种策略：一种策略是插入“侦察刺激”，即在实验过程中，随机地插入几次只给“预备”信号但不呈现刺激的实验操作，以观察被试是否会在未接收到刺激时作出反应；另一种策略是不断地改变刺激间的时间间隔，或者不断地改变在每次给予被试“预备”信号到呈现刺激之间的预备时间(一般在 1.5 秒到 2 秒范围内变化为宜)，这样可以避免被试形成反应定势。

第二，要选择的反应与要辨别的刺激一一对应

在选择反应时间测定中，要选择的反应的数量和要辨别的刺激的数量相等，二者有一一对应关系。为什么一定要相等呢？如果选择反应时间的测试装置上可以呈现红光和绿光两种刺激，但却只有一个反应键，然后要求被试实验中必须先看清刺激光的颜色时才能按键。这样的装置就是有缺陷的，因为即使被试能自觉地按照主试的要求参加实验，但实验中总有一种尽快反应的倾向，就很容易使实验变成了简单反应时间的测试，或者被试会因为担心过早反应而使反应滞后。这一装置必须再增加一个反应键，然后要求被试看到什么颜色的灯光刺激，就按与之对应的那个键，如果按错就不会停止计时，被试还必须尽快纠错，这就会使测得的反应时间反而加长，所以利用这样的装置进行选择反应时间测定就会有效地控制被试“抢码”的次数。

第三，使用计算机程序控制反应时间测定时要注意计算机的配置和设置

国内许多心理学实验室配备了反应时间的计算机测试程序，同时有许多研究者在进行认知心理学研究时要使用计算机程序控制实验过程，并用计算机记录反应时间。在这些实验中，要尽量做到让所有被试使用的计算机的配置和设置保持一致。计算机的配置和设置一致，主要是指计算机的 CPU(中央处理器)速度、内存大小、显示器的大小、显示器的分辨率和刷新速度、键盘响应速度、安装和启动的操作系统等等，这些因素或多或少地与时间记录的准确性有关。

如果利用 E-Prime 实验软件编写反应时间测试程序，最好使用与该软件配套的反应盒。所以在购买 E-Prime 软件时，最好同时购买相应的反应盒。欲了解更详细的相关信息，可登录网站：www. E-Prime. com。

二、影响反应时间的刺激因素

测得的反应时间与刺激条件、主体因素均存在密切联系。刺激条件主要是

指刺激信号本身的特点和呈现刺激的环境因素，如刺激的性质、刺激的强度、刺激的作用时间、刺激呈现的背景等；主体因素主要包括被测者的年龄、性别、知识经验或教育训练、动机因素、人格特征及身心状态等。这里首先介绍影响反应时间的刺激因素的有关研究。

（一）反应时间的通道效应

反应时间的通道效应就是指反应时间因刺激的感觉通道不同而不同，表5-1概括了许多研究结果得到的与各不同感觉通道对应的简单反应时间。①从表5-1的数据看出，不同感觉通道对应的简单反应时间差异很大，其中的原因是多方面的，如刺激的方式与能量性质的不同；感觉器官接受刺激的方式、能量转换的方式和传递方式等的不同；刺激信息对机体来说其生物学意义大小的不同等等。就视觉与听觉通道来说，刺激作用到引起神经冲动需要的潜伏期存在明显差异。使用猫进行研究的结果表明其视觉刺激存在较长的潜伏期。使用声音直接刺激猫耳，可以在1～2毫秒引起神经冲动，从发出声音信号到猫竖起耳朵只需要8～9毫秒的时距；用光刺激猫的眼睛，其引起神经冲动并传递到大脑视觉皮层则需要20～24毫秒；如果绕过视网膜，用电极直接刺激视神经，则神经冲动可以在2～5毫秒达到视觉皮层。可见，信息在视觉神经通路上的接收、转换和传递消耗了较多的时间，这是视觉通道反应时间长于听觉通道反应时间的主要原因。一般，光线经过眼球折光之后直接刺激视网膜，但在视网膜上不能立即引起神经冲动，而是需要完成光化学反应这一中介过程才行。其他通道间的差异可能各有其不同的原因。

表5-1　简单反应时间的感觉通道效应

感觉通道	简单反应时间(ms)	感觉通道	简单反应时间(ms)
触觉	117～182	温觉	180～240
听觉	120～182	嗅觉	210～390
视觉	150～225	痛觉	400～1 000
冷觉	150～230	味觉	308～1 082

当然，同一感觉通道中，刺激的性质不同（如味觉刺激中的酸、甜、苦、辣刺激）、刺激位置不同（如视觉刺激作用于网膜的不同点），也会造成反应时间不同。

① 朱滢：《实验心理学》，北京大学出版社2000年版，第148—149页。

还有研究表明，刺激的不同组合作用引起的反应时间也有所不同，比如声音加电击刺激的反应时间比单独使用声音刺激或电击刺激的反应时间都要短，光、声、电击三种刺激同时作用时反应时间更短。

(二) 刺激强度与反应时间的关系

刺激强度通常反映的是刺激的能量大小，一般情况下刺激能量越大，反应越快，反应时间越短；刺激能量越小，反应越慢，反应时间越长。研究表明，刺激强度与反应时间的变化不是线性关系，如表 5-2 和图 5-6 所示。

表 5-2　对不同强度的 1 000 Hz 纯音刺激的反应时间

lg*I*	*RT*/ms	lg*I*	*RT*/ms	lg*I*	*RT*/ms
0	402	1	193	6	124
0.2	316	2	161	7	118
0.4	281	3	148	8	112
0.6	249	4	139	9	111
0.8	281	5	130	10	110

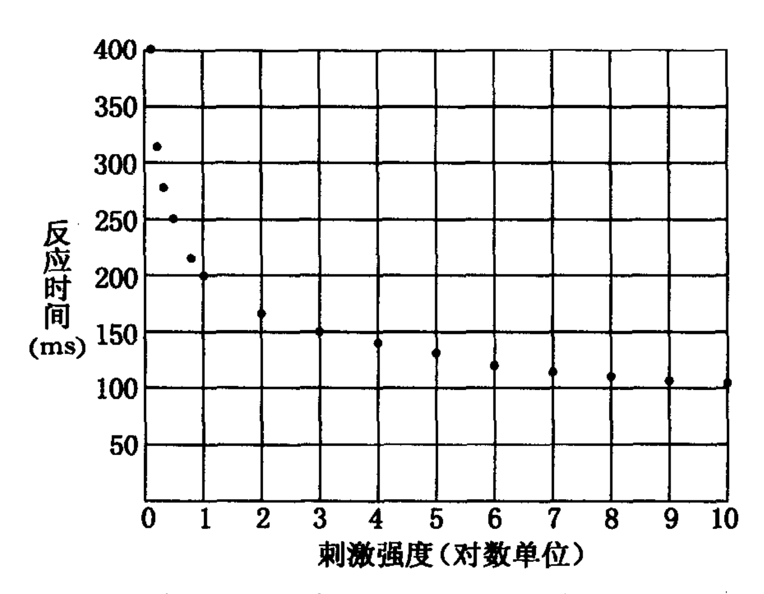


图 5-6　对不同强度的 1 000 Hz 纯音刺激的反应时间

将表 5-2 的数据表示成图 5-6 的形式，可清晰地看出，从较弱的 1 000 Hz 纯音开始，随着刺激强度的增加，反应时间缩短得很快。但随着刺激强度的继续增加，反应时间缩短的速度越来越慢，似乎在逐渐接近一个极限值。具体地说，在 $\lg I=0$ 时，表示刺激强度接近于阈限强度，对应的反应时间为 402 ms；当 $\lg I=1$ 时，刺激强度大约是阈限强度的 10 倍，反应时间缩短到 193 ms，约为阈限强度时

的一半；当 lg$I$＝2 时，刺激强度又增加了 10 倍，反应时间缩短到 161 ms，与 193 ms相比只缩短了 30 ms。lg$I$ 取值在 8～10 范围内变化时，反应时间几乎没有发生什么变化。

（三）刺激呈现的时空特征对反应时间的影响

刺激呈现的时空特征主要包括刺激呈现的不同位置、面积大小及不同时距等。比如，当刺激强度不变、刺激呈现时距一定时，增加视觉刺激的面积，由于感觉器官的累积效应，也会增加刺激的心理强度，进而影响到反应时间，表 5-3 所示的实验结果说明了这一点。

表 5-3　方形视觉刺激的面积与反应时间的关系

方形视觉刺激的边长(mm)	3	6	12	24	48
反应时间(ms)	195	188	184	182	179

表 5-3 的结果说明，随着刺激面积的增加，反应时间逐次减少，但是面积累积效应是有限的。当面积增加到一定程度时，反应时间就难以继续缩短，累积效应就不明显了。研究也发现，空间累积效应也表现在双眼视觉和双耳听觉上。在一个实验中，对同一个光刺激，双眼观察时的反应时间比单眼观察时的反应时间更短。在听觉方面也有类似现象，同样强度的声音作用于单耳的反应时间为 147 毫秒，而作用于双耳时的反应时间为 133 毫秒。

此外，改变刺激的时间特性也会引起个体反应时间变化。当刺激的物理强度不变时，增加刺激作用于感官的时间，同样可造成刺激的累积效应，即时间累积效应，刺激引起的心理强度也随之增加。弗罗伯格(Froeberg, 1907)曾做过一个关于光作用于眼睛的久暂对反应时间影响的研究，其实验结果如表 5-4 所示。①

表 5-4　光刺激的久暂与反应时间的关系

光刺激的持续时间(ms)	3	6	12	24	48
反应时间(ms)	191	189	187	184	184

表 5-4 的实验结果显示，刺激持续的时间越长，反应时间越短，但是这种变化关系也有一定的限度，如表中数据显示，当刺激作用时间从 24 毫秒增加到 48 毫秒时，反应时间并没有随之缩短。

① 转引自：孟庆茂、常建华：《实验心理学》，北京师范大学出版社 1999 年版，第 146 页。

(四) 刺激复杂度对反应时间的影响

反应时间不仅受到某些刺激因素的单独作用,而且会受到刺激因素组合的作用,包括刺激材料复杂度的影响,这主要是指对辨别反应时间和选择反应时间的影响等。显然,对于辨别反应时间和选择反应时间来说,刺激的数量越多,复杂度越高;刺激的可分辨性越差,复杂度越高。刺激的复杂度越高,其反应时间应该越长,梅克尔(Merkel, 1885)的早期研究直接说明了这一点①。梅克尔要求被试对阿拉伯数字的1～5分别用右手的5个手指作反应,对罗马数字的Ⅰ～Ⅴ分别用左手的5个手指作反应。整个研究包括很多个实验序列,每一实验序列中使用的刺激选择数目各不相同,分别为1～10。在一个实验序列开始前都会告知选择刺激的数目,实验结果如表5-5所示。

表5-5 可供选择的刺激数目与反应时间的关系

可供选择的刺激的数目(mm)	1	2	3	4	5	6	7	8	9	10
选择反应时间(ms)	187	316	364	343	487	532	570	603	619	622

亨蒙(Henmon)的实验则能说明刺激辨别难度的变化对反应时间的影响。方法是:在被试的前面一左一右呈现两种不同颜色的刺激,主试指定的颜色出现在哪一边哪一边手就作反应。例如,当主试指定被试要对红光作反应时,就让红色与其他各种颜色配对呈现,红光出现在哪一边就用哪一边的手作反应。实验结果显示刺激之间的差别越小,辨别反应或选择反应时间就越长,如表5-6所示。②

表5-6 刺激辨别难度与反应时间的关系

要辨别的刺激	平均反应时间(ms)	要辨别的刺激	平均反应时间(ms)
白和黑	197	红和黄	217
红和绿	203	红和橙	247

三、影响反应时间的机体因素

除刺激因素外,反应时间也与一系列机体因素有关,包括被试的适应与准备

① 转引自:孟庆茂、常建华:《实验心理学》,北京师范大学出版社1999年版,第148页。
② 张庆云:《基础实验心理学》,河南大学出版社1993年版,第163页。

状态、年龄与教育训练、动机水平、个性差异和实验时的心向等。

(一) 被试的适应与准备状态

当被试对外界环境的适应水平不同时，其反应时间会有所不同。例如，被试在光适应和暗适应的不同水平上进行实验：让被试在250勒克斯(lx)照度条件下对一个距离眼睛约30.48 cm的白色圆盘作反应。如果实验前被试所在环境的光照度条件不同，则反应时间也有所不同，如表5-7所示。实验前被试的光照环境光线越强，被试的适应水平越低，其反应越慢，即反应时间越长；反之，实验前环境光线越弱，被试适应水平越高，反应越快，即反应时间越短。这一规律也适合其他感觉通道，而且因此可以将反应时间作为个体适应状态的一个测量指标。

表5-7　反应时间与被试对光的适应水平的关系

被试实验前所在环境的光照度(lx)	200	150	100	50	0
反应时间(ms)	154	146	144	140	131

反应时除与被试的适应状态有关外，也与其准备状态有关。在简单反应时间实验中，从预备信号发出到刺激呈现之间的时间间隔叫预备时间，预备时间的长短会影响到反应时间。一般来说，预备时间过长或过短都会使被试的反应变慢，即反应时间延长。预备时间过长，会使被试在注意力高度紧张后很快处于心理疲劳或松懈状态，反应时间会加长；如果预备时间过短，被试还来不及做好反应前的准备，也会造成反应时间延长。许多研究表明，预备时间在1.5秒左右最有利于被试的反应。有人记录运动员起跑的反应时间，将从预备口令到鸣枪声的时间间隔定为1秒、1.5秒和2秒三种时间，结果如表5-8所示，1.5秒预备时间时反应时间最短。①

表5-8　预备时间与被试的反应时间

时间(s)	反应时平均数(ms)	标准差
1.0	198.92	19.750
1.5	170.75	11.348
2.0	201.87	18.214

① 张庆云:《基础实验心理学》,河南大学出版社1993年版,第160页。

(二) 年龄与教育训练

在人的整个发展过程中的 25 岁以前,随着年龄的增长,个体反应越来越快,反应时间逐渐缩短,此后一直到 60 岁左右则是反应时间相对稳定的时期,通常变化不大。60 岁后,人的反应逐渐变得迟缓,反应时间会随着年龄增长逐渐延长,这是机体衰老的表现。

当然,在人的发展中,教育训练也会促进反应速度的提高。许多体育项目都可以促进个体反应速度的提高,如武术、球类、击剑等。对于特定的反应任务来说,短时间的练习也会缩短反应时间,不过教育训练和练习对反应速度的促进毕竟是有限的,当其达到一定水平后就难以继续提高。

(三) 动机因素

在反应时间实验中,可以采用不同的奖惩措施引起被试的额外动机,以观察不同动机水平下被试的反应时间。如利用被试在反应时间实验中,一般都喜欢自己的反应又快又好的心态,对被试的反应进行赏罚。"赏"是每次反应以后及时告诉被试反应的结果,罚是当被试反应慢于某一水平时利用反应给予电击。这两种条件与不给任何诱因的试验随机地交替进行,借以抵消练习效应,其实验结果如图 5-7 所示。

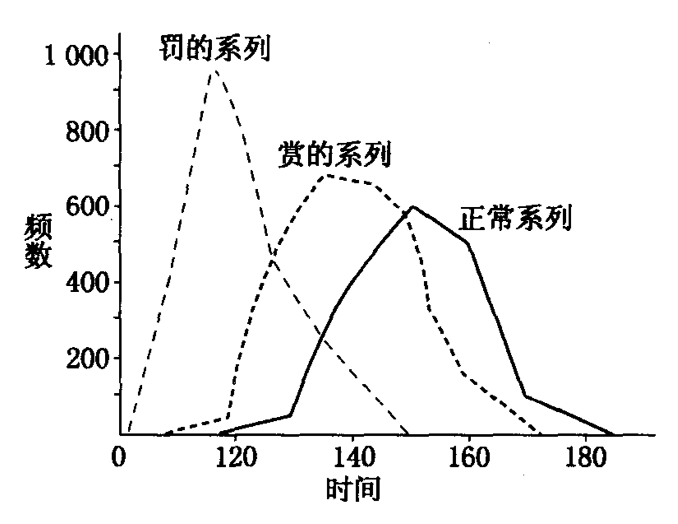


图 5-7　附加动机对反应时间的影响

(四) 个体差异

个体差异作为被试变量,也是被试难以控制的,这种变量表现在反应时间上,不仅个体之间存在差异,也表现在同一个体在不同条件下,由于心理和生理状态的变化,反应时间也有所不同。研究认为,反应时间的个体差异也反映了个

体神经活动类型的不同，一般可将个体的神经活动类型分为三种类型：感觉型、运动型和平衡型。感觉型的人在反应时间实验中，会将注意力更多集中在刺激的感知上，刺激出现时需要较多时间将注意中心转移到运动发动方面，所以反应时间往往较长。运动型的人在反应时间实验中则会把注意力更多集中在运动发动上，所以刺激一旦出现就会更快地发动反应，但是由于分配较少的注意力给刺激感知，所以对刺激的判断容易出现错误，即容易出现“抢码”或“抢跑”的现象。平衡型的人则比较好地把注意力分配给刺激感知和反应发动上，既能较快地发动反应，也较少出现“抢码”的情况。

(五) 被试心向：速度与正确率权衡

在信息加工心理学中，常常把反应时间看作对信息的心理加工过程复杂程度的指标，根据反应时间的延长推断信息加工阶段的增加。但反应时间受多方面主客观因素影响，除前文讨论的一系列因素外，它还明显受到被试心向的影响，即追求高速度还是追求高准确性。在反应时间实验中，被试有时会以牺牲反应速度为代价去换取高的反应准确率，有时会以牺牲准确率为代价去换取快的反应速度。在不同的实验要求和实验条件下，被试会综合考虑多种利弊，建立一个权衡反应速度与反应准确率的标准来指导他的反应，这就是反应速度与准确率的权衡现象(speed-accuracy trade-off，简称 SAT)。

一般而言，被试可以在准备充分的情况下作出反应，也可以在准备不充分的情况下作出反应。在充分准备下，因其能提取全部与刺激有关的信息而作出反应，此时，准确率总是很高，速度却慢些。在准备不充分的情况下，即在没有足够时间或不允许被试提取全部与刺激有关的信息的条件下，被试还可利用在加工阶段初期积累的部分信息去作出反应，也可以利用某种不同于上述完全加工的信息加工方式(如猜测加工)去作出反应。而对这种不完全的信息加工过程的研究，可以获得信息加工最初阶段的加工特性。

在反应时间实验和 SAT 现象的研究过程中，研究者们利用 SAT 现象形成了一种不同于传统反应时间的新的反应时间实验范式，即速度与准确率权衡范式(SAT 范式)。

在 SAT 实验中，可以通过控制被试作出反应的时间长短作为实验自变量，测定在各种时间条件下的反应准确率，然后以反应时间为横坐标，以反应准确率为纵坐标，可得到一条速度与准确率权衡曲线。图 5-8 是一条用反应信号法进行再认实验的 SAT 曲线。在这条曲线上，可以找到在各种反应时间下的反应准

确率(d')。在这条曲线上,可以用3个参数来描述SAT现象。I表示截距,R表示准确率(d')随反应时间(RT)的变化而变化的速率,A表示准确率的渐进值,即在无限延长提取时间条件下的最高准确率。

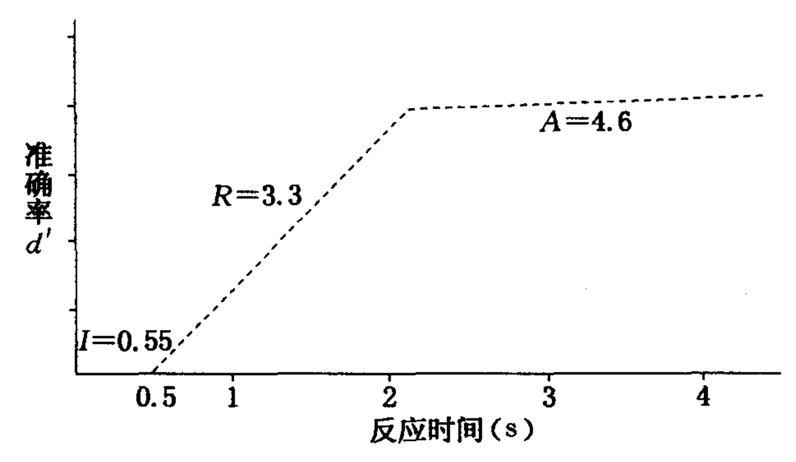


图 5-8　再认提取实验的 SAT 曲线(Vickelgren et al., 1977)

图5-8显示,在信息加工的早期,反应时间与准确率之间的权衡关系是非常显著的。当接近或达到A值之后,随着反应时间的延长,准确率的变化很小。

第三节　反应时间测量的应用

由于反应时间是在某种任务情境中对刺激的快速应答过程中测量的,它能够有效地反映机体发展的水平、身心状态和完成任务时信息加工的复杂性,所以在心理学研究、实际应用中有着特殊地位和价值。在此,我们从三方面讨论反应时间测量的实际应用。

一、人才选拔中的反应时间测量

由于反应时间是个别差异的重要方面,它不仅能间接反映个体内部的信息加工能力、技能操作水平,而且能够直接影响到许多领域的工作绩效,对个体在特定任务情境下的工作绩效具有一定的预测功能,因此反应时间测试成为人才选拔工作中一项常规的测试指标。就国内情况看,人才选拔中使用反应时间测量的领域主要包括:航空航天招飞(要求在复杂而快速多变条件下迅速决策与操作)、运动员选拔(短跑运动员起跑要快,但又不能经常"抢码",击剑运动员、球类

运动员要有较短的“追踪反应时间”和“综合反应时间”等，详见下文介绍)、公务员选拔(要思维敏捷)等。这些领域一般都希望选拔出具有“思维敏捷”、“反应灵活，动作协调”、“处事大胆果断”、“反应速度快”等心理品质的人员，而这几方面的心理品质都可以直接或间接地用反应时间来测试。比如“思维敏捷”可以通过“抢答”反应时间、联想反应时间等来测试；“反应灵活，动作协调”可以通过复杂反应时间来测试；“反应速度快”可以通过简单反应时间来测试。具体测试的手段很多，有时需要一些专门的仪器设备，有时也可以“就地取材”。这里介绍两种专门的反应时间测试方法：言语(联想)反应时间和复杂反应时间。

(一) 言语(联想)反应时间的测试

言语反应时间，也叫做联想反应时间，是指通过视觉或听觉通道给被试呈现一个语词刺激并同时开始计时，被试接收到刺激后尽快地作自由联想或控制联想，并通过音键(麦克风)报告联想到的语词，计时装置即停止计时。这一过程记录的时间即为联想反应时间或言语反应时间。

联想方式包括自由联想和控制性联想两大类。自由联想就是不对被试联想的方向施加任何限制，要求被试在接收到刺激词语后，尽快报告其最先想到的那个词语。“自由联想能够反映被试思维的敏捷性和灵活性。一般来说，智力高的反应时间短；智力低的反应时间长。”①控制联想则对被试的联想方向提出明确要求，被试接收到刺激词语后尽快报告联想到的符合要求的词语，通常会要求被试回答刺激词语的同义词、近义词、反义词，或有对比关系的词语等。具体的测试材料和测试程序一般包括：

材料与呈现：词单与呈现程序。在实验前为被试准备好要其作联想反应的词单，同时设计好词语呈现的方法和程序。呈现方法和程序的设计要依据实验条件而定，比如可以由主试读给被试听(或播放录音给被试听)，可以在光电速示仪上逐个地呈现给被试看，可以利用计算机程序控制在显示器上呈现或通过耳机呈现等。

被试的反应操作与计时：根据材料呈现方式，被试按要求完成对词语刺激的联想反应，并记录反应时间。

实验设备与操作：较为传统的方法是通过速示仪呈现刺激，通过音键(麦克风)来记录被试的反应和反应时间，即刺激呈现计时装置开始计时，被试对着音

① 杨博民：《心理实验纲要》，北京大学出版社 1989 年版，第 401 页。

键作出语音反应即停止计时，这样就很容易记录反应时间，这里都是利用被试报告的语音信息控制计时。也可以利用录音或摄像手段记录时间，比如将听觉的刺激呈现和被试的言语反应全程录下来，然后利用语音分析软件来计算刺激语音与反应语音之间的时距，这一方法简便易行，只需要一只采样率较高的录音笔和较好的语音分析软件就可以了。[①] 还可以使用计算机程序控制实验的进程和记录实验数据。

实验结果的分析：对于联想实验的结果，除对反应时间进行计算和分析之外，还应对反应的内容进行分析，这不仅包括对反应词语的质量进行分析，也包括对反应词语中表现出来的知识范围、情绪状态等方面进行分析。

（二）复杂反应时间的测试

上文介绍了简单反应时间、辨别反应时间和选择反应时间的概念及测试方法，这其中辨别反应时间和选择反应时间属于复杂反应时间范畴。在现实生活和工作中，属于复杂反应时间范畴的例子还有很多，除辨别反应时间和选择反应时间外，还有如追踪反应时间、综合反应时间等，如打球、击剑等都需要较好的综合反应速度。这里介绍辨别反应时间和综合反应时间的测试方法。

追踪反应时间，就是当刺激数量不止一个时，要求被试作快速跟踪反应，即当某一个刺激出现时被试就尽快作出与该刺激相对应的反应，而被试事先并不知道哪一个刺激会出现。追踪反应时间与选择反应时间比较接近，但又有所不同，我们结合图 5-9 来加以说明。

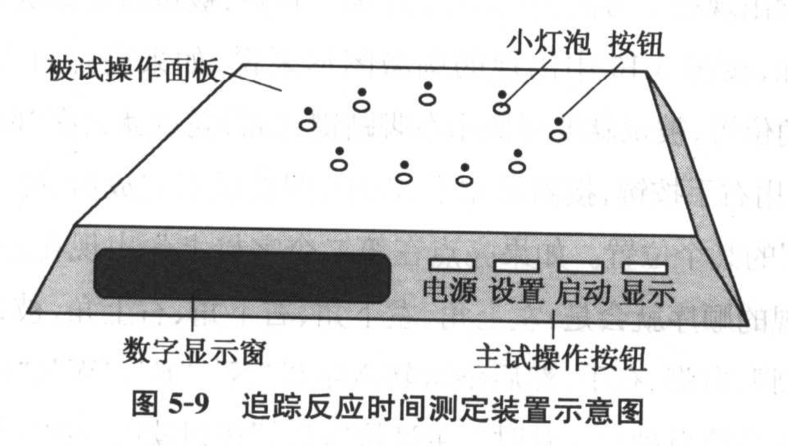


图 5-9　追踪反应时间测定装置示意图

如图 5-9 所示是笔者根据传统的“注意分配测试仪”设想的追踪反应时间测定装置。在被试操作面板上有九个小灯泡，每个小灯泡的旁边都有一个按钮。

① 南京师范大学建立了“语图分析实验室”，可以对语音材料进行时长的分析。

主试的操作面板包括电源开关、实验次数设定按钮、启动键、测试结果显示按钮和测试结果的数字显示窗。实验中,主试接通电源后,先设定实验次数,然后告诉被试作跟踪灯泡的反应,即看到哪一个灯泡亮就尽快按该灯泡旁边的那个按钮,连续做下去直到仪器发出"嘀"的结束音。被试清楚操作方法后,主试发出"预备"口令,然后按启动键即开始实验,直到完成设定的实验次数,仪器自动停止计时。主试通过"显示"按钮可以从显示窗读取数据:设定的实验次数、实验所用的总时间、错误次数、正确反应的平均反应时间等。

如果没有专门仪器,该反应时间可以利用注意分配测试仪中的视觉刺激部分作为替代设备。

综合反应时间测试就更为复杂,它要考虑到被试手脚同时参与的反应模式,其测试装置的简化模型如图 5-10 所示。该装置包括两部分并以通讯线连接,一部分是主试操作面板,主要用于主试对实验的控制和读取实验结果;另一部分主要是给被试呈现刺激,这些刺激是一系列字母型图形,每个图形中都有上下左右四个顶点,分别对应于四个被试反应键:左手键、右手键、左脚键、右脚键。实验开始时,主试先向被试交代实验任务和操作方法,打开仪器电源开关,然后按1～8号图形选择按钮中的某一按钮,即选定了呈现给被试的刺激图形。此时,刺激屏幕的左上角有一个亮点,要求被试看着这一亮点做好准备,主试喊"预备"口令后按"启动"键,实验开始,实验装置同时开始计时。实验开始后,亮点会按照一定顺序连续出现在字母的顶点,出现在哪个顶点,被试就要尽快用手或脚按对应的键。比如,按图 5-10 中出现的刺激图形来说,如果亮点首先出现在字母"M"的右下角位置,被试就要尽快用右脚踏键,随后亮点就会在"M"的右上角出现,被试尽快用右手按键,接着是左手、左脚,四次反应完成后,亮点会出现在第二个字母"U"的某个位置。如果亮点在第二个字母上先出现在左上角,则其在该字母上出现的顺序就会是:左上角、左下角、右下角、右上角,被试反应的顺序就是左手、左脚、右脚、右手,然后依次转入字母"N"、"E"、"W"、"S",完成对"S"的反应后实验装置自动停止计时。主试按"显示"按钮读取该次实验中被试反应的总时间、错误次数、平均反应时间等实验结果,再按"复位"即回到待机状态。

这一反应时间测试中,需要被试快速作出反应,而要达到快速反应,被试必须具有良好的手脚协调性。我们把这一反应时间叫做综合反应时间,它是许多实践领域都有要求的一项指标。

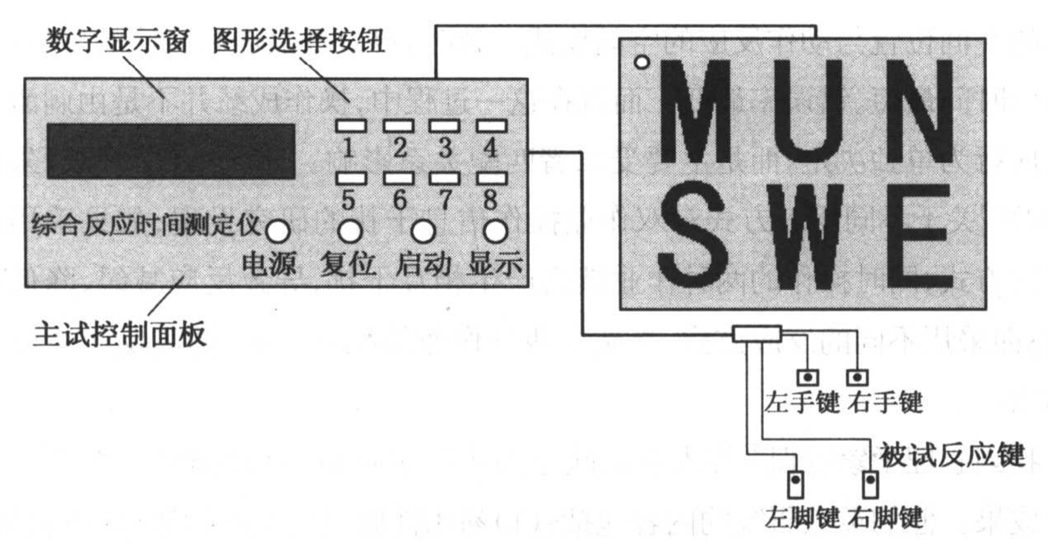


图 5-10　综合反应时间测定装置模式图

二、工程设计中的反应时间测量

反应时间测量也属于工程心理学(engineering psychology)或人因工程学(ergrgonomics)研究中的基本技术,比如利用测量反应时间对人机系统的设计进行评估,根据在特定条件下人的反应速度的有限性进行人机系统设计等等。这里,我们对国内外工程心理学研究中的反应时测量问题进行简单介绍,比如刺激-反应相容性研究。

刺激-反应相容性(stimulus-response compatibility,简称 SRC)是人在对刺激进行加工过程中表现出来的一种现象,它是工程心理研究的最核心问题(Alluisi & Warm, 1990)。① 可以认为,当一定的刺激和反应匹配会产生较好较快的结果时,这样的刺激-反应匹配就具有了相容性。② 很明显,刺激-反应相容性是工程设计中应追求的重要目标。

刺激-反应相容性研究源于第二次世界大战美国空军对显示屏的研究需要,但真正使其成为心理学研究课题的是菲茨等的研究。③菲茨和西格(Fitts & Seeger, 1953)系统地研究了空间刺激与动作响应之间的相容性问题,结果发现④:当

① Alluisi, E. A. & Warm, J. S. (1999). Things that go together, stimulus-response compatibility. In R. W. Proctor (Ed.), *Stimulus-response compatibility: an integrated perspective*. North-Holland, Elsevier Science Publisher, BV, 33—86.

② 刘艳芳:《S—R 相容性:概念、分类、理论假设及应用》,《心理科学》,1996,19(2):105—109。

③ 张学民、舒华:《实验心理学纲要》,北京师范大学出版社 2004 年版,第 201—202 页。

④ Fitts, P. M. & Seeger, C. M. (1953). S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46:193—210.

刺激的空间位置与动作反应的空间模式一致时，可获得最佳和最快的反应结果，即反应时间最短、错误率最低。而且在这一过程中，操作成绩并不是由刺激模式或反应行为单独决定，而是主要受二者匹配关系影响。还比如，葛列众、朱祖祥(1992)①关于不同反应方式对双作业操作信息干扰的研究发现，如果采用相同的反应方式，同时操作的两种作业就会产生相互干扰，导致反应减缓、降低作业绩效；而采用不同的反应方式，将减少两种作业的相互干扰，提高反应速度和作业绩效。

我国心理学家张侃②等人也在此方面进行了卓有成效的研究，得到一系列重要成果。他们重点研究的内容包括：(1)易用性原则与空间刺激-反应相容性；(2)可懂性原则与语义刺激-反应相容性；(3)重要性、功能和使用与群体模板相容性；(4)信息冗余原则与刺激-刺激相容性、反应-反应相容性；(5)警告信号与线索相容性等。这方面的研究对于人机系统的设计来说意义重大，许多人机系统就是一个刺激-反应的操作系统，其从设计方面能否实现刺激呈现与反应操作的最佳匹配直接关系到人机系统是否能满足刺激-反应相容性的要求，也就直接关系到其是否能在使用中获得最佳效果。因此，人机系统设计过程中要在具体情境中进行大量试验，以寻找最佳的设计方案。在对各种设计方案进行评估和筛选时，反应时间是其中最重要的衡量指标。

工业设计中反应时间的应用当然非常广泛，不只是刺激-反应相匹配的问题，还有人因限制性问题。在特定系统中，人受到生理和心理资源的限制，其反应速度有限。虽然可以通过人机系统匹配性的提高来提高操作者的反应速度，但这种改进和提高必然有限。我们要研究在各种特定情境中人的反应速度极限，以便在系统设计和参数设置时防止操作人员的负荷超载。

三、交通领域的反应时间测量

交通领域中的反应时间因素主要包括三个方面：道路设计，包括道路的宽窄、曲直、交通信号及各种路标的配置；驾驶员的疲劳测试与告警；驾驶员的测试

---

① 葛列众、朱祖祥：《不同反应方式对双作业操作信息干扰的影响》，《心理科学》，1992，21(3)：226—330。

② 张侃：《刺激-反应相容性原理与工程心理学研究》，见中国心理学会编：《当代中国心理学》，人民教育出版社 2001 年版，第 513—519 页。

与培训等。

(一) 道路设计与各种信标配置

从道路设计来说,必须考虑人的反应速度的有限性。首先,要根据人的反应速度来设计交叉路口及相关信号,比如在驾驶员到达交叉路口时,要能提前看清交通信号及相交的其他路段的交通状况,包括行人情况和来往车况,提前做好必要的准备。在一些道路状况较差路段,要设置警示标志和限速标志,而且这些警示标志的位置要前移,即驾驶员可以提前看到。其次,在高速公路设计中,要注意直行路段不能过长,因为在直行路段,驾驶员的警觉水平下降,遇到紧急情况时其反应时间会较长;最后,要对高速公路的行车距离加以限制和提示,减少追尾等交通事故的发生频率。

我国心理学工作者詹美莎曾利用反应时间的测量对道路标志的易懂性进行了研究。① 有 40 名 20～52 岁的职业汽车司机参加了实验。实验材料是将未达到理解度标准的 15 种交通标志图示制成带有各种街景、广告及人物等背景的幻灯片,共三套,每套 15 张,如图 5-11 所示。每种标志在各幻灯片画面上出现的位置是随机的,三套幻灯片中每种标志在画面上的位置又各不相同。用幻灯机将刺激物呈现于 30×20 平方厘米的屏幕上。用数字计时器及自动控制装置记录被试回答标志图示的意义的声音反应时间。被试坐椅距离屏幕 2 米,话筒放在坐椅前的桌子上。在主试向被试说明本实验的目的在于测定 15 种标志的辨

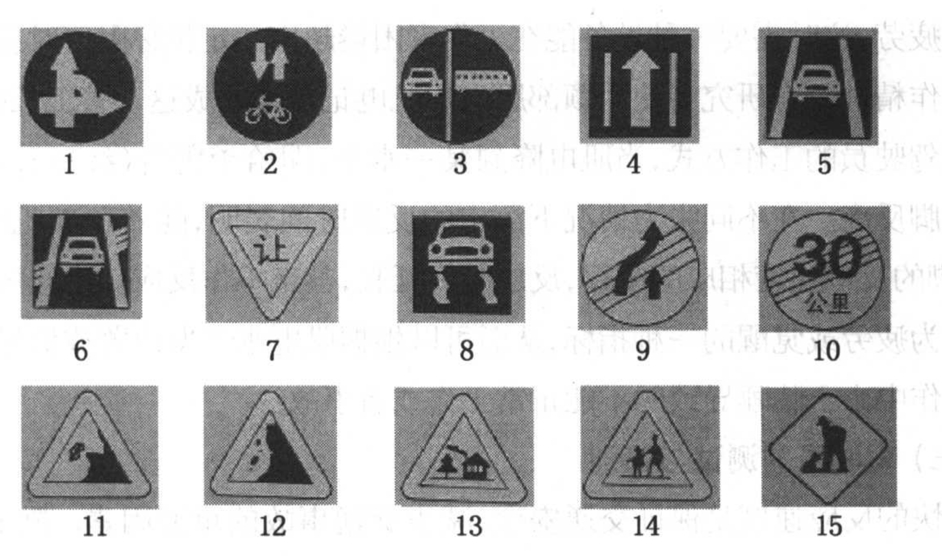


图 5-11 各种交通标志物示意图(詹美莎,1987)

① 詹美莎:《北京市道路交通标志的心理学评价》,《心理学报》,1987,19(2):167—174。

认时间后，接着将15种幻灯片依次呈现给被试并解释每种标志的意义及用途。然后，要求被试在听到主试预备口令后注视屏幕，当画面出现时迅速从中寻找交通标志并回答其意义，尽量做到又快又准。

结果发现：(1)除9号"解除禁止超车"标志，其他标志的再认正确率均超过85%，9号标志的再认正确率最低，为82.5%；(2)各标志平均反应时间差异显著，其中9号标志反应时间最长，为2.38秒，4号(单行路)、6号(快速路终点)、11号(山旁险路)标志的反应时间也较长，均超过2秒；5号(快速路)、14号(学校)、15号(施工)、3号(只准两种车通行)、8号(准试煞车)和13号(村镇)标志的反应时间都比较短，在1.48～1.75秒之间。这些实验结果告诉我们，不同路标的可懂度是不一样的。

除对其可懂度进行实验研究外，我们还可以对路标设计的大小、颜色、背景、安置的位置等开展研究，研究中可以将反应时间作为重要的反应变量。这样的研究可以为改善路标设置提供非常直接的、有价值的数据。

（二）疲劳测试与告警

疲劳情况下的反应速度也会下降，有相当多的交通事故就是由于疲劳驾驶造成的，所以有必要设计出相应的检测设备，通过检测驾驶员的反应时间来评定其疲劳水平，以确定其是否还能继续驾驶车辆。

肌肉的紧张度与反应时间有关，肌肉的紧张程度可以作为疲劳的指标。①日常生活中，当一个人昏昏欲睡或丧失警觉时，若有一个东西使他警觉起来是非常必要的。如司机长时间驾驶车辆、监测人员长时间监测某种机器或仪表都容易引起疲劳，这时需要一种设备能在工作者困倦超出一定指标限度时发出警告，使人振作精神。有研究者使用颈部肌肉的肌电记录来完成这一警卫任务。实验时模拟驾驶员的工作方式，当肌电降到某一水平，即给予警告(红灯)，司机必须作出抬脚反应。在不同张力情况下测出的反应时间表明，随着颈部肌电张力的降低，脚的反应速度相应的变慢，反应时间延长，甚至不作反应。由此可见，肌电可以作为疲劳或觉醒的一种指标，从而可以根据肌电水平发出警戒信号，预防在某些操作中由于瞌睡导致的不能正常工作或者事故。

（三）驾驶员的测试与培训

较快的反应速度是保证交通安全，减少交通事故的重要因素。随着社会经济发展，我国城镇私家车数量迅速增长，"驾驶员"的数量相应迅速增加，有条件

① 张庆云：《基础实验心理学》，河南大学出版社1993年版，第164—165页。

买车的就可以成为"驾驶员"。根据前文论述,反应时间与个体的神经活动类型和个性特征关系密切,有些个体的反应速度较慢,遇到紧急情况时甚至会出现应激性"休克",不能采取有效措施处理应激事件。所以,可以说,有些个体不适合驾驶车辆。目前,我国驾照考试还缺少必要的心理测试成分,使得一部分"很不适合"开车的人成为驾照持有者,这一部分人成为公路上的潜在"危险者"。

我国汽车销量猛增,使汽车驾驶员的培训也成为非常重要的任务。培训中,应该增加对这些人应激反应的训练,缩短其反应时间,以提高其紧急情况下的反应速度。当然,反应速度的提高也不能仅从应激训练方面来做,还需全面提高驾驶员的整体心理素质,包括责任心。比如,培养其不酒后驾车的自觉性等,因为"酒精在脑神经系统达到一定浓度时,中枢神经系统逐渐迟钝,对周围情况变化的反应速度大大下降"。①

阅读材料 5-1

酒精及药物对反应时间的影响

各类酒中都含有不等量的酒精。酒精在脑神经系统达到一定浓度时,中枢神经系统逐渐迟钝,对周围情况变化的反应速度大大下降。如果酗酒,其反应时间将延长 2～3 倍,甚至更长,往往紧急情况已到眼前,还未发现或未采取任何措施便已肇事。例如车速每小时 40 公里,未饮酒的驾驶员对道路复杂情况作出反应只需 0.6 秒,而饮酒后的驾驶员却要 1.8 秒。这样,汽车在反应时间内行驶的距离就从 0.6 秒的 7 米增加到 1.8 秒的 21 米,所需的安全距离增加了 14 米,即从 27 米增加到 41 米。由于酒精对大脑皮层的抑制过程会产生破坏作用,使驾驶员很难正确估算车速、距离和控制自己的能力,以致驾驶员动作不准确、不适当。据测试,当血液中的酒精达到 0.5～0.7 毫克时,驾驶员不仅选择反应时间会增长,而且错误反应概率也会增加 46%。

目前,刺激中枢神经的心理杀伤性药剂,大致分三类,即镇静剂、兴奋剂和致幻剂。镇静剂虽能消除情绪紧张、焦虑、恐惧感,但会使驾驶员肌肉活动能

① 杨治良:《实验心理学》,浙江教育出版社 1998 年版,第 140 页。

力下降，并出现睡意，以致反应迟钝。兴奋剂对中枢神经系统的作用与镇静剂作用相反，它可以消除疲劳、驱逐睡意、改善思维活动力、提高反应速度，但会使各类职业人员思想麻痹，过高估计自己的能力。致幻剂会使人有时产生幻觉。如伤风感冒服用过量解热镇痛剂，会使人的注意力、精力、反应能力下降。其他诸如安眠药、麻醉药品、止咳药、止痛片、咖啡因、非那明、利血平等类药品，要遵医嘱，不应随意服用。另外有些工程人员不应服用抗胺类药品，因为它会使人昏昏欲睡，反应能力下降。因病住院，或在治疗过程中，最好主动介绍自己的职业，以及近期是否有出差等任务，以便医生合理给药。还有一种情况，即不能靠喝浓茶、浓咖啡来增加兴奋，因为这种兴奋是短暂的，随之而来的则是疲劳、困倦。

——资料来源：《实验心理学》（杨治良，1998）①

阅读材料 5-2

常见的反应时间法研究范式

启动研究范式是近几十年来基础心理学研究领域最常用的范式之一，研究者运用前置线索为启动刺激，对目标识别与判断的认知过程进行研究，可揭示信息加工的内在机制。认知心理学兴起后，启动研究范式作为一种重要的研究方法在感知、注意、语言认知、意识的研究等领域得到了广泛应用。下面介绍一些基于启动范式发展起来的一些具体研究方法。

1. 移动窗口范式

移动窗口范式（moving windows paradigm）是阅读研究中常用的一种技术，该范式实验的基本程序是让被试自定步速，逐字（或词）地阅读一段文字，句子材料以字（或词）为阅读单元逐一呈现，被试每按一次鼠标，当前词便消失，同时下一个新的词出现。在实验过程中，计算机将自动记录被试每次按键的时间，并自动计算被试连续两次按键动作的时间间隔，并将它作为被试对每个字、词或句子阅读时间的指标。采用该技术可以测量被试加工句子的不同结构单元所用的反

① 杨治良：《实验心理学》，浙江教育出版社 1998 年版。

应时间，并据此分析阅读理解的规律。

2. 注意线索范式

注意线索范式(attentional cueing paradigm)是选择注意研究领域常用的方法，该方法的基本程序是在注意目标呈现之前，给被试呈现不同性质、特征或有效性的线索刺激，通过不同线索对注意目标的提示作用，探讨被试对注意目标进行注意加工的规律。通常采用该范式研究选择注意时，选择的线索特征包括线索的有效性、线索呈现的位置(如外周还是中央)、线索与目标特征的一致性、线索有效性的概率等等。

3. 注意追踪范式

注意追踪的研究范式是多目标追踪范式(multiple object tracking paradigm，简称MOT范式)。该研究范式的基本程序是，让被试追踪一系列由不同特征组成的、变化的、随机运动或按照特定的规律运动的目标(target)，当目标发生特定变化时，被试的注意也会发生变化，此时要求被试根据实验要求对被追踪目标的变化作出不同反应，并记录其反应时间和错误率，该研究范式被很多研究者应用于注意追踪的实验研究。

在部分追踪研究中，研究者加入了目标融合(target merging)技术，即通过把两个刺激融合在一起(如两个"圆"用一条线连起来，成为"哑铃")改变追踪对象的呈现形式，通过测量这种不同形式的多目标融合的反应时间，可以发现人对不同空间组织和空间关系的若干物体的注意追踪与加工机制，该研究方法在近几年的注意追踪研究中经常被研究者采用。

4. 空间线索技术

空间线索技术(space cueing paradigm)是研究刺激-反应相容性的一种方法，该方法的基本程序是在给被试呈现不同复杂程度的目标之前，先给被试呈现一个与目标特征具有一致性或部分一致的线索刺激，通过不同相似性的线索刺激对目标的提示作用，来考察被试目标刺激的加工和识别速度。

其他还有很多基于反应时间的研究技术，如通过语音记录反应时间的DM-DX技术以及其他的心理学实验设计软件，通过眼动仪和计算机记录被试眼动过程与视觉停留时间的眼动技术等等。

——资料来源：《实验心理学纲要》(张学民，舒华，2004)[①]

① 张学民、舒华：《实验心理学纲要》，北京师范大学出版社2004年版，第189—190页。

建议阅读文献

1. 朱滢主编:《实验心理学》,北京大学出版社 2000 年版,第 125—160,276—288 页。
2. 杨治良编著:《实验心理学》,浙江教育出版社 1998 年版,第 105—161 页。

复习思考题

1. 如何理解反应时间、简单反应时间、选择反应时间、辨别反应时间、减法反应时间法、加因素法、开窗实验、速度-准确性权衡?
2. 如何测定简单反应时间、选择反应时间、辨别反应时间?
3. 简述反应时间研究的历史。
4. 认知心理学研究中的反应时间方法主要有哪些?
5. 举例说明减法反应时间法的基本原理。
6. 反应时间测量中需要注意的问题主要有哪些?
7. 加因素法实验的基本假设是什么?
8. 影响反应时间的因素有哪些?
9. 反应时间测量在实际中的应用主要有哪些?

第六章
感觉的实验研究

本章内容提要

感觉是有机体接受外部刺激并为知觉加工提供信息资源的过程，是感性认识过程的初级阶段，也因此成为心理学基础实验研究的重要领域之一。感觉包括外部感觉和内部感觉两大类，其中视觉和听觉更受关注，因为人类活动中需要的关于外部世界的大部分信息都是通过视觉和听觉接收的。本章仅就视觉和听觉现象的刺激测量、感官的基本结构、信息加工机制和典型实验进行分析。具体内容包括：(1)眼睛的结构与视觉刺激测量，主要介绍眼睛的结构及视觉信息的传递、光度测量；(2)视觉现象及其实验分析，主要介绍颜色视觉的形成机制与基本现象、暗适应的实验研究及其应用、视敏度的测量及影响因素等；(3)耳的结构与听觉测量，主要介绍听觉的适宜刺激、听觉信息的接收、传递与测量；(4)听觉现象及其实验分析，主要介绍音高测量及其与声强的关系、响度测量及其与声频的关系、听觉掩蔽的实验分析、听觉告警的应用及设计要求等。

感觉(sensation)是哲学心理学和实验心理学的早期研究课题之一，也是心理学中最基础的实验研究课题。人类长期受到物质现象与精神现象之间关系的困扰，原因之一就是我们无法确认物质与精神是不是有直接的对应关系。有时我们看到物质对精神的决定性，有时又看到二者完全相反的关系。为解决这样的困惑，研究者自然想到要探查外部信息进入精神领域的第一个环节即感觉过程，只有把握这一环节的特征和机制，我们才可能对物质与精神的关系有基本认知。在传统实验心理学中，感觉阈限的测量成为其典型范式，因限于当时研究的技术水平，感觉的内部机制研究未能获得很大的进展。在现代实验心理学中，引入神经生理技术之后，我们已经可以在皮层水平、细胞水平甚至分子水平上对感觉进行分析。作为基础性研究，我们还需要考察关于感

觉量的基本测定方法。

感觉的种类很多，根据感觉刺激是来自有机体外部还是有机体内部，可以把感觉划分为外部感觉(external sensation)和内部感觉(internal sensation)。外部感觉包括视觉(vision)、听觉(audition)、嗅觉(olfaction)、味觉(taste)和肤觉(skin sensation)等五种，内部感觉包括机体觉(organic sensation)、运动觉(motion sensation)和平衡觉(balance sensation)等三种。在人类的各种感觉中，视觉和听觉最为重要，占据主导地位，因此本章以视觉和听觉为例来分析感觉的基本神经机制及实验现象。

第一节 眼睛的结构与视觉测量

视觉是人类获取外部信息的主要通道，在人的感觉中占主导地位。要理解视觉机制及其表现出来的各种现象，需要分析视觉器官的结构特征及对应的适宜刺激的特征。

一、眼睛的结构及视觉信息传递

(一) 视觉的适宜刺激

视觉的适宜刺激是可见光(visible light)。光是一种电磁波(electromagnetic wave)，而电磁波具有一个巨大的波谱范围，其中只有一小部分是能够被人和动物的眼睛觉察到的，这一部分就叫做可见光谱或可见光。对人来说，可见光的频谱范围大约在波长为380毫微米至780毫微米①之间。如图6-1所示，这一可见光谱范围大约只占整个电磁频谱的1/70。另外，在可见光谱的邻近区，有红外线(infrared ray)和紫外线(ultraviolet ray)，红外线的波长大于780毫微米、紫外线的波长小于380毫微米。

可见光具有三维物理特征：波长(wavelength)、振幅(swing)和纯度(purity)，与此相对应的三维心理特征是色调(hue)、明度(brightness)和饱和度(saturation)。在可见光谱上，波长不同的光，会引起不同的色觉。如果眼睛接收到的刺激光的波长范围在380毫微米至420毫微米之间，就会有紫色光的感觉印象；如

---

① 毫微米(nm)也就是现在使用越来越频繁的“纳米”，1纳米$=10^{-9}$米。

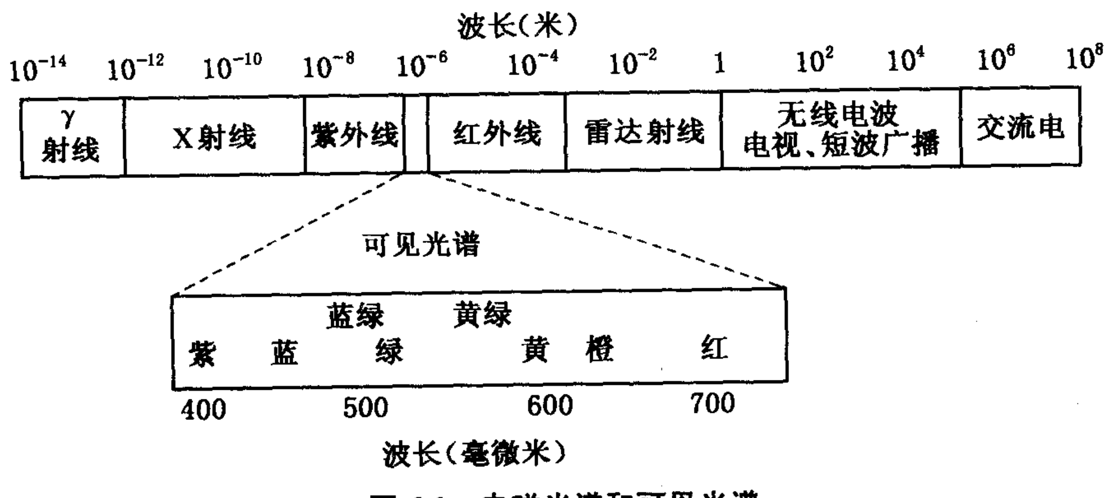


图 6-1 电磁光谱和可见光谱

果眼睛接收到的刺激光的波长范围在 630 毫微米至 780 毫微米之间,就会有红色光的感觉印象。表 6-1 列出了能引起不同颜色感觉的光谱的主波长及其范围。①振幅是光的强度或能量单位,它与视觉的明度有关。纯度是指光波成分的复杂程度,它与颜色视觉的饱和度有关。颜色视觉的饱和度取决于刺激光线中优势波长所占的比例,决定色调的优势波长的光所占比例越大,该色调的饱和度就越高,反之则饱和度越低。例如,紫红色、粉红色和浅绿色等都是饱和度较低的彩色,而真红色和鲜绿色则是饱和度较高的彩色。

表 6-1 刺激光的波长范围与视觉颜色的对应关系(单位:毫微米)

感觉到的颜色	主波长	波长范围
红(red)	700	630—780
橙(orange)	620	600—630
黄(yellow)	580	565—590
绿(green)	550	500—570
青(cyan)	500	470—500
蓝(blue)	470	420—470
紫(violet)	420	380—420

光是一种物理刺激,但颜色感觉是一种心理现象,可是,这种物理刺激为什么会引起颜色感觉呢?要理解这一点,就要查明眼睛的结构及视觉神经传递机制。

① 关于不同颜色视觉的刺激光的波长范围,各种教科书的记录很不一致,表 6-1 中的数据是综合了各种文献后给出的一组波长取值范围,也只是一种参考值,不能作为标准值。

(二) 眼睛的结构与机能

人的眼睛是一个前后直径约为23毫米的近似球状体，由眼球壁和眼球构成，大体结构如图6-2所示。眼球由三层膜状结构及内容物构成，其膜状结构由三层组成：最外层叫做巩膜，白色，所以俗称“眼白”，主要起巩固、保护眼球的作用，它只有前端部分是透明的，叫角膜，这是外部光线进入眼球并聚焦的最初入口；中间层是富含黑色素的脉络膜，其中充满血管，担负着为眼睛输送营养的任务，而且脉络膜中的黑色素可以保持眼球内部的黑暗并吸收来自其他方向的散射光，脉络膜的前方是虹膜，虹膜的中心开口就是瞳孔，瞳孔由虹膜肌(包括缩瞳肌和扩瞳肌)控制，可以放大或缩小，相当于照相机的光圈，随着外部光线的强弱而收缩或放大，从而调节进入眼球的光线量；眼球膜状结构的最内层是视网膜，或者叫网膜，网膜的中心正对着瞳孔中心，叫做中央窝，网膜上散布着两种感光细胞，负责将经由眼球传递的光刺激能量转换为神经能，视网膜神经汇聚成束离开眼球与人脑连接，在其成束离开眼球的地方出现一个无神经细胞的小区，叫做视神经乳突或盲点。

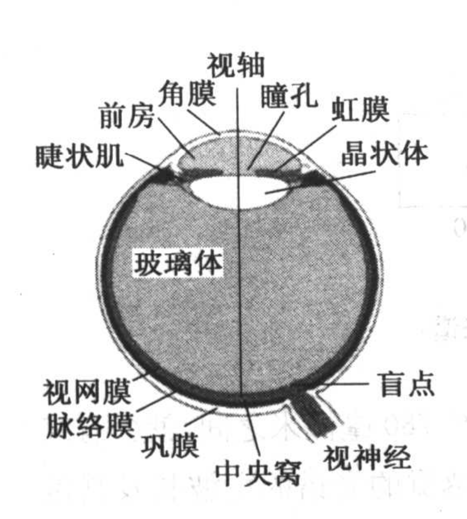


图6-2 人类左眼球的剖面图

眼球的主要内容物包括晶状体、前房水和玻璃体三部分。晶状体把眼睛分成大小不等的两半——小的叫前房，大的叫玻璃体。前房充满液体，可用于维持眼球的形状；玻璃体也充满透明液体，它可以保持晶状体的位置并使它具有弹性。晶状体附近的睫状肌可以控制其形状，即睫状肌可根据接收的刺激光线的入射角度调节晶状体曲率，以保证把光线聚焦到视网膜上形成清晰的像。从上述介绍可以看出，在眼球结构中，角膜和晶状体是其屈光系统的主要构成部件。

(三) 网膜结构与视觉信息传递

视网膜(retina)的厚度一般在0.1～0.5毫米，它是一种多层的网状结构，在视网膜的不同部位，其网状结构也不一致。视网膜由三层神经细胞组成，最外层(靠近脉络膜)是大量的感光细胞，包括视锥细胞(cone cell)和视杆细胞(rod cell)两类。视锥细胞和视杆细胞分布的区域、数量、结构形状和机能都有所不同。视锥细胞密集地分布于视网膜中央窝的大约1平方毫米的区域，数量为600多万个，呈锥状，负责检测亮度、颜色、物体细节等信息；视杆细胞分布于中

央窝以外的较广大区域，数量大约为1.2亿个，呈棒状，在暗适应条件下，视杆细胞能够对微弱的光刺激发生反应。视网膜的第二层结构主要是双极神经元，这些双极神经元与外层的感光细胞相连接，其数量大约是感光细胞的1/100，这就意味着有许多感光细胞与一个双极神经元相连接，在这里必定发生了信息的整合。但是，视锥细胞与双极神经元几乎是一对一的关系，所以每个视锥细胞获得的信息可以直接向内部传递，这也是中央窝区域能够对刺激物的空间大小作精细反应的原因所在。由于有很多视杆细胞与一个双极神经元连接，所以在微弱光线下，每个视杆细胞接受到的刺激能量可在双极神经元得到放大。视网膜的第三层结构（靠近玻璃体）是神经节细胞，它与双极神经元相连接，但它的数量又远远少于双极神经元，所以在这一连接环节也必定会发生信息整合。神经节细胞的轴突在眼球的内层汇集成束，这就是视神经纤维。视神经纤维又汇聚成神经盘（即视神经乳突），然后穿出眼球，将视觉信息传递到中枢神经系统。

从人的左、右眼球后面发出的视神经形成“半交叉”，即部分交叉：来自鼻侧视网膜的视神经纤维都交叉行至对侧，其传递的是颞侧视野的光刺激信息；来自颞侧视网膜的视神经纤维则不交叉，其传递的是鼻侧视野的光刺激信息，这种半交叉传递机制可使得左视野中的光刺激信息传递到大脑右半球，而右视野中的光刺激信息传递到大脑左半球。由视神经经过视交叉后组成的视束的一部分纤维终止于中脑上丘，参与瞳孔对光反射和视觉运动反射等活动；而大部分视束纤维终止于外侧膝状体，由视觉传导通路的第四级神经元接替。

在正常情况下，人眼对光的强度具有极高的敏感性，感觉阈限很低。特别是波长500纳米左右的光比其他波长的光更容易被觉察到。光刺激离中央窝8～12度视角时，视觉感受性最高。一般认为，刺激盲点，其对光没有感受性（不过，新近有研究发现，刺激盲点的强度达到一定水平时，盲点也有微弱的感受①）。

二、光度测量单位

视觉研究中，经常要进行光度测量，但光度测量的单位种类繁多，容易引起混乱。表6-2给出的是一种国际单位制中的常用光度单位，我们提倡在研究中

① Michael, S. G., Richard B. I. & George, R. M. (Eds.)(2002). *Cognitive neuroscience: the biology of the mind*. Norton, 184.

使用统一的国际单位制光度单位,以便于研究结果的相互比较,以及研究者的相互认同。

表 6-2 国际单位制(SIE 单位制)的光度单位

几何学	光度量			
	名称	符号	单位	英文名
光源	光强度	I	坎德拉	candela(cd)
点光源 立体角 A	光照度	E	勒克斯	lux(lx)
B	光亮度	L	尼特	nit
C	光通量	ϕ	流明	lumen(lm)

[转引自杨治良(1998):《实验心理学》,第 275 页]

(一) 光强度

光强度(luminous intensity)一般是对于发光光源来说的,是指光源的发光功率,以单位时间、单位立体角内发出的光能量为测量标准,单位为坎德拉(candela,简称 cd)。

(二) 光照度

光照度(illuminance)是指单位时间从光源照射到单位面积上的光能量,单位为勒克斯(lux,简称 lx)。对于一个点光源来说,光照度与光源的发光强度成正比,与到光源的距离的平方成反比。

(三) 光亮度

光亮度(luminance)是指一个物体表面的明亮程度,它取决于从物体表面反射出来的光能量,也就是说它取决于物体表面接受到的光照度和物体表面的光反射率两个方面。不同物体对光有不同的反射系数或吸收系数,所以同样强度的光照射到不同的物体表面,其反射光的强度有所不同,得到的光亮度也有所不同。如在同样的光照条件下,白纸看上去显得很亮,黑纸则显得很暗,这就是因

为它们具有不同的光反射率：常用白纸的光反射率大约为20%，黑纸的光反射率大约为3%。

光亮度的单位较多，最常用或最直接的是尼特(nit)(坎德拉/平方米，1 nit=1 cd/m²)。其实，光亮度可以与光照度使用相同的单位。

(四) 光通量

光通量(luminous flus)是单位时间内由光源向某一立体角或某一面积上辐射的光能量，也就是在单位时间内穿过某一立体角或某一面积的光能总量，其单位为流明(lumen，简称 lm)。

从上述四个基本光度单位的概念，可以看出它们之间的关系：

第一，光通量与光强度、光照度的关系。对于光源来说，其向某一立体角内或某一面积上辐射的光通量等于光源强度与光的辐射角度(立体角)或面积之积，即 $\phi = I \times A$ 或 $I = \phi/A$，其中 A 代表光辐射的立体角度或面积大小；对于接受光照的物体表面来说，物体表面接受到的光通量等于光照面积与光照度之积，即 $\phi = E \times A$ 或 $E = \phi/A$，其中 A 代表光照面积的大小。

第二，光照度与光亮度的关系。光照度与光亮度都是对于一个物体表面来说的，但前者描述物体表面接受光照的情况，也就是单位面积上的入射光通量；后者描述物体表面反射光的情况，也就是单位面积上的反射光通量，二者的函数关系为 $L = R \times E$。其中，R 为物体表面的光反射系数，它是指物体表面对入射光的反射比率，即 $R = (L/E) \times 100\%$。

在实际的光度测量中，有物理测量和心理测量两种方式。物理测量使用各种标准光度计进行直接或间接测量，其中最常用的是照度计。有了照度计，不仅可直接测量光照度，也可以间接测量光强度、光通量、光亮度及物体表面的光反射率等。心理测量则是将已知标准光源(如国际标准烛光)与待测光源进行视觉比较，这是一种较粗略的间接测量方法。

第二节　视觉现象及其实验分析

一、颜色视觉及其实验分析

在视觉现象中，最神奇和最令人感到不可思议的莫过于颜色现象，它使我们能够看到一个“绚丽多彩”的世界！以至于几乎每个人都无法相信：物理世界本

没有颜色;颜色只不过是一种感觉现象、心理现象,并不是一种物理现象;不同的颜色只不过是不同波长的可见光在人脑中的映象不同而已!因此,我们现在最想知道的就是,颜色视觉的生理机制是什么。

(一) 颜色视觉理论

颜色视觉理论,或色觉理论,就是关于颜色视觉发生机制的理论,这些理论几乎都是立足于从光谱进入视觉系统后的神经生理过程来分析的。色觉理论主要包括:扬(Young)的三色说、赫尔姆霍兹(Helmholtz)的三色说、柯尼格(König)的优势调制说、莱德-富兰克林(Ladd-Franklin)的三色说、黑林(Hering)的色颉颃说、亚当斯(Adams)的区域和阶段说、缪勒(Müller)的区域和阶段说等。这些理论可以概括为两大类:"一类是扬-赫尔姆霍兹(Young-Helmholtz)的三色说;另一类是黑林的色颉颃说。至于像缪勒等的区域和阶段说则只是前两种学说的综合。"①

1. 扬-赫尔姆霍兹三色说

三色理论(trichromatic receptor theory/Young-Helmholtz theory)最早由英国科学家托马斯·扬(Thomas Young)于1802年提出来,德国生理学家赫尔姆霍兹在1852年注意到这一理论,随之进行了研究,于1860年发表《光学》一书对该理论加以扩展并使之逐渐为人们所接受。② 赫尔姆霍兹假定,视觉系统只需三种感受体就可以产生所有的颜色。他设想在视网膜上存在三种视觉细胞,即α、β和γ细胞,这三种细胞各自包含具有完全不同光谱敏感特性的色素,三种色素的敏感光谱分别在450 nm、530 nm和650 nm附近,分别对应于蓝色光、绿色光和红色光光谱区域③,如图6-3所示。按照这一理论,如果刺激光主要是短波段光谱,容易被α视锥细胞中的蓝色素吸收从而激活此类视觉细胞,形成蓝色感觉;同样,中波段光谱会形成绿色感觉、长波段光谱会形成红色感觉。如果某一光谱同时被红色素和绿色素吸收,则这两类细胞都能在一定程度上被激活,于是形成介乎红色与绿色之间的颜色感觉,即形成黄色感觉,这就是为什么580 nm的光谱会形成黄色感觉的原因(如图6-3)。近代的研究基本上证实了扬-赫尔姆霍兹的假设。马克斯(Marks)等人用显微光谱光度方法研究了人和灵长类动物

---

① 俞文钊:《实验心理学》,浙江教育出版社1989年版,第232页。

② 林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社2003年版,第1517页。

③ Atkinson, R. C., et al. (Eds.)(1988). *Stevens' handbook of experimental psychology*. New York: John Wiley & Sons, Inc., 205.

的视锥细胞。他们在显微镜下把单色仪发出的非常细微的单色光束投射到某单个视锥细胞上，然后测量各单色光穿过视觉细胞的量，以确定该细胞的光谱吸收特性。①可以认为，某一波长的光被视锥细胞吸收的越多，该细胞内包含的物质对这一光谱越敏感，这种物质可以被称为是与该光谱对应的色素。这一精细的实验研究发现，网膜细胞中，主要存在视锥细胞中的三种主要色素，这三种色素的吸收峰值分别在 445 nm、535 nm 和 570 nm 附近，并且具有较宽范围的光谱感受性。这三种色素分别对短波、中波和长波光谱敏感，这在一定程度上证实了赫尔姆霍兹(1856)的假设，也验证了视锥细胞中包含三种主要色素，即蓝色素、绿色素和红色素。不过，按照马克斯的研究，红色素的敏感光谱在 570 nm，这在可见光谱中应属于绿光光谱，这一点让人有些困惑，尚需作进一步的精细研究。

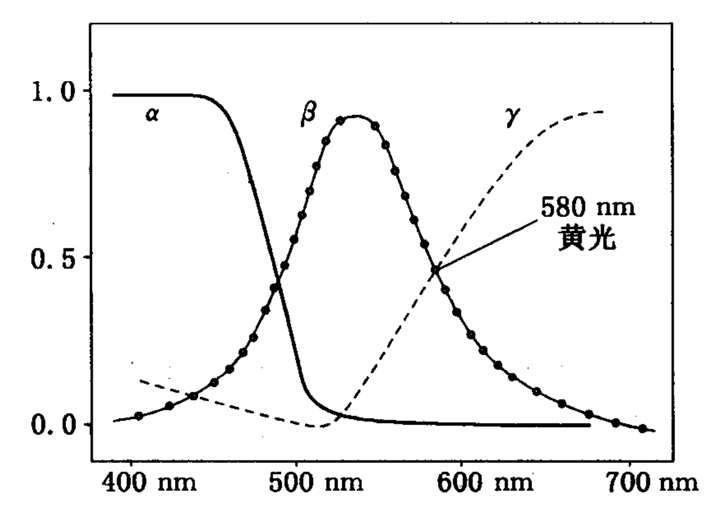


图 6-3 赫尔姆霍兹(Helmholtz, 1856)假设的视神经细胞光谱敏感曲线 (Atkinson, 1988)

2. 颉颃说

颉颃说(opponent-process theory)，也叫做“对抗过程理论”、“颜色视觉颉颃理论”(opponent theory of color vision)或“四色说”，最早由德国心理学家黑林于1864年提出。这一理论建立在颜色的互补或对抗这一事实基础之上，认为存在红、绿、黄、蓝四种基色，再加上黑、白两种，就有六种基本感觉。假定视觉神经系统存在红绿感受器、黄蓝感受器和黑白感受器，每种感受器在受到色光刺激时，能发生对抗互补作用。颉颃说能够较好地解释颜色的互补现象和后效现象，而

① 转引自：朱滢：《实验心理学》，北京大学出版社 2000 年版，第 191 页。

且也得到有关实验结果的支持。20世纪60年代,美国心理学家用显微光谱光度计对视网膜和视觉神经通路进行实验时,发现了三类神经节细胞:一类细胞对所有可见光谱都反应,负责报告明度信息;一类细胞对红光发生正电位反应,对绿光发生负电位反应;一类细胞对黄光发生正电位反应,对蓝光发生负电位反应。这些细胞能够估量一类相反颜色的相对强度,因此被称为对立细胞或颉颃细胞。[①]就是说,神经节细胞中存在三对颉颃细胞:红-绿、黄-蓝、白-黑(对各种波长的可见光都有反应,反映明度信息)。

3. *颜色阶段理论*

三色说和颉颃说使用不同的神经过程说明颜色视觉的形成机制,都在一定程度上成功地解释了颜色视觉的某些现象。随着后续大量神经生理实验研究结果的发表,上述两种理论也都得到一定程度的证实。但是,越来越多的证据也表明,颜色视觉并不是某单一阶段神经过程的结果,而是多个阶段不同神经过程的共同结果,因此缪勒等人提出颜色视觉的两阶段理论,这一理论实际上就是三色说与四色说的综合。该理论认为,视网膜视锥细胞接收到光刺激时,首先引起对应的光敏色素的反应并激活相应的视锥细胞,而三类视锥细胞激活的相对强度信息传递到下一阶段的神经节细胞。三组对抗神经节细胞按照视锥细胞的激活强度编码颜色信息,这其中包括强度信息(反映在明度视觉方面)、色调信息,而颉颃细胞对颜色信息进行编码之后将其输送到大脑相应皮层,并在皮层激活或抑制某些中枢过程,进而发生颜色感觉。

概括地说,颜色视觉是一系列神经过程的结果。红、绿、蓝三种基本颜色首先在视网膜视锥细胞中以不同色素的光化学反应的形式得到加工,然后颜色信息在视觉通路的不同水平上以兴奋-抑制的对抗过程逐步得到编码。[②]

(二)颜色视觉的特征

由于不同光谱的刺激光经过视觉系统的加工后会形成各种不同的颜色视觉印象,所以视觉也叫做颜色视觉或色觉,而颜色视觉又分为彩色与非彩色两大系列。非彩色系列是指从黑色到白色及其间深浅不同的灰色,基本特征是明度变化。就物体表面来说,从黑色到白色的系列中,其光反射率相差很大,接近白的

---

① Atkinson, R. C., et al. (Eds.)(1988). *Stevens' handbook of experimental psychology*. John Wiley & Sons, Inc., 1512.

② Schiffman, H. R. (1996). *Sensation and perception*. John Wiley & Sons, Inc.

一端,光反射率可达到80%以上;接近黑的一端,光反射率只有10%以下。从黑到白,明度逐渐增加。

彩色系列除具有明度变化的特征外,还具有另外两方面的基本特征,即色调、饱和度变化。为便于理解这三方面特征的关系,心理学家制成了如图6-4所示的颜色锥体图。①颜色锥体(又称光锥)的纵轴代表白、灰、黑系列的明度变化:顶端是白色,底端是黑色,中间是灰色。纵轴上的每一点均代表不同亮度的非彩色,由全白到全黑,而其饱和度不变。圆周代表不同的色调:红、橙、黄、绿、蓝和紫等色。锥体上的每一点,均代表一种颜色,而锥体的全部,则代表所有不同颜色的集合。锥体的水平截面均为圆形,从中间向上和向下移动,截面圆的半径都逐渐减小。截面圆的半径代表了对应颜色的饱和度。换句话说,锥体各点到纵轴的垂直距离代表颜色饱和度的变化,与纵轴垂直距离越短,饱和度越小;与纵轴垂直距离越长,饱和度就越大。

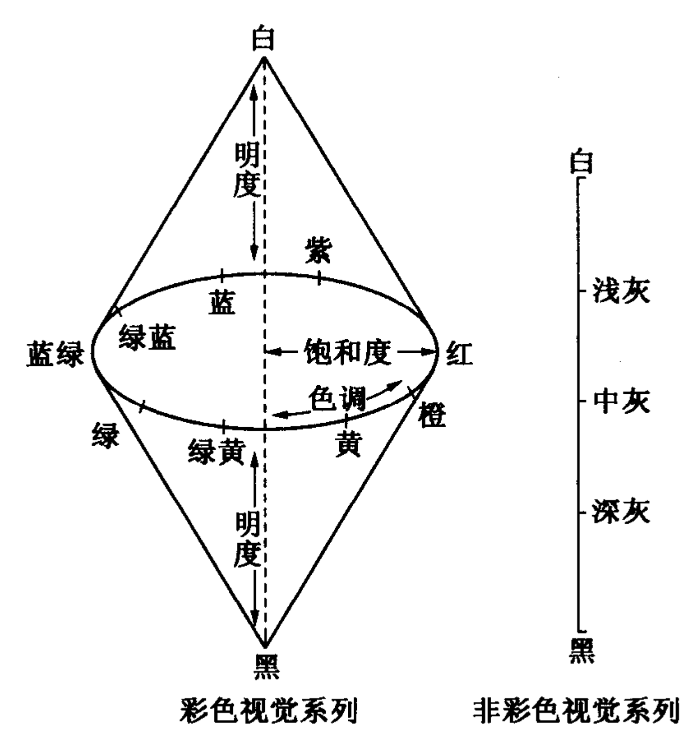


图6-4 颜色视觉锥体图:彩色和非彩色系列

关于颜色视觉,还需要区分的是:光的颜色与颜料的颜色。光的颜色,是由直射到视网膜上的主要光谱的波长直接决定的,而颜料的颜色则是由于颜料本身吸收了照射光中的某些光谱后剩余反射光达到眼睛的视觉印象。有时我们看

① 赫葆源、张厚粲、陈舒永:《实验心理学》,北京大学出版社1983年版,第538页。

到，同一个彩色物体在不同光照背景下颜色视觉效果完全不一样，这是因为不同条件下的照射光光谱含量不一样，被其吸收后剩余的反射光谱含量也就不一样了，因此看上去颜色就有所不同。

（三）颜色混合与CIE系统

在日常生活中，进入我们眼睛的很少是单色光，而是包含多种光谱成分的混合光；我们看到的绘画颜料、物体表面的颜色也往往是多种颜料混合而成的。简单地说，颜色视觉常常是颜色混合的结果，而颜色混合又包括色光混合和颜料混合两种。

1. 色光混合

色光混合就是把不同颜色或不同波长的光同时照射到视网膜的同一部位，从而产生一种新的颜色视觉的过程。由于色光混合是将不同的光加在一起同时刺激视网膜，所以也叫做颜色相加法。混合后的刺激强度大于相加前的刺激光，色光混合会增加颜色视觉的明度。大量实验发现，整个可见光谱色中的每一种颜色均可由红光、绿光和蓝光按各种不同的比例混合得到，红、绿、蓝就被称为色光混合的三原色或基色，而且当三基色按照大致相等的比例混合在一起时可以得到白光。

红色＋绿色＋蓝色＝白色

色光混合还有补色律和间色律之分。如果两种颜色的光按适当比例混合可得到白光，这两种色光称为互补色；当两种色光混合时，即使调整比例也不能得到白光时，就只能得到中间色，这就是间色律。实验发现了以下混合方程式：

间色律混合方程式：红色＋绿色＝黄色　红色＋蓝色＝紫色　蓝色＋绿色＝青色
补色律混合方程式：黄色＋蓝色＝白色　红色＋青色＝白色　紫色＋绿色＝白色

青、紫、黄分别是三基色红、绿、蓝的对应补色。每一对互补的色光都可按照适当的比例混合得到白光，如红光和青光、绿光和紫光、蓝光和黄光混合都可以得到白光。彩色电视机或电脑显示器主要利用了加色法的基本原理，荧光屏上的色光组合都是红、绿、蓝三基色相加混合的结果。

2. 颜料混合

色光混合是一种加色法，但是在日常生活中，我们见到的颜料、油漆等物质，如果按照不同颜色混合，得出的颜色与色光混合得到的颜色不一样，如黄色颜料和蓝色颜料混合得到的是绿色颜料而不是白色颜料，这是为什么呢？原来，颜

料、油漆类的混合配色遵循的是减色法。一般来说，颜料的颜色是这种颜料吸收白光中一定波长的光谱后剩余光谱的色调，如黄色颜料能够吸收入射白光中的蓝光而反射红光和绿光，反射出来的红光和绿光同时投射到人眼而混合在一起产生黄色感觉，因此我们称这种颜料为黄颜料；青色颜料能够吸收入射白光中的红光而反射蓝光和绿光，蓝光和绿光同时投射到眼睛而混合在一起产生青色感觉；那么，如果将黄颜料和青颜料混合在一起，情况会怎样呢？这样混合后的颜料中实际上就包含有原来的黄颜料和青颜料了，其中的黄颜料会从入射白光中吸收蓝光，青颜料会从入射白光中吸收红光，于是入射光被反射的成分主要是绿光，因此我们看这种混合后的颜料便得到了绿色感觉。减色法的三基色是黄、青、紫，即黄颜料、青颜料和紫颜料混合在一起得到的是黑颜料，因为三种颜料各自吸收入射光中的一种基色，入射白光就几乎全部被混合后的颜料所吸收，我们就看到混合后的颜料为黑色。

减色法颜色混合的三基色正好是加色法颜色混合的三基色的补色。彩色电影胶片的画面则是由黄、青、品红三种影片染料按减色法处理构成的。减色法颜色混合的方程式如：

黄色＝白色－蓝色
青色＝白色－红色
紫色＝白色－绿色
黄色＋紫色＝白色－蓝色－绿色＝红色
紫色＋青色＝白色－绿色－红色＝蓝色
黄色＋青色＝白色－蓝色－红色＝绿色
黄色＋紫色＋青色＝白色－蓝色－绿色－红色＝黑色

这里需要注意两点：第一，减色法颜色混合总是会降低反射到眼睛的刺激光强，所以减色法颜色混合后的明度会降低；第二，减色法颜色混合方程式中等号的左边均指颜料的颜色，右边则指相减的色光(黑色除外)的颜色。

在心理学实验室里，可以使用简易方法演示颜色混合。颜色混合的方法很多，最常用的颜色混合实验方法是使用电动混色轮(色轮)。这是一个由不同颜色扇形色纸组成的圆盘，套在旋转器的轴上，在转速达到闪光融合临界频率(约 30 转/秒)时，即可产生一种均匀的混合色。混合色的比例取决于每种被混合的色纸的大小比例。但是，混色轮实验具有以下问题：(1)产生的颜色很不饱和；(2)由于

多次使用，色盘磨损严重，边缘处有翘边现象；(3)混色轮运转时，发出“嗡嗡”声，噪声较大。另外注意，使用混色轮演示的颜色混合实际上是一种颜料混合。

另一种颜色混合的方法是采用一套已知透光率的良好滤色片，使用透光率不同的滤色片可以得到光谱中的各种单色光。随后，把它们投射到不反光的白色屏幕或人眼视网膜的同一部位上，这种方法比较方便和精确，是一种色光混合。

为了解决传统实验过程中的种种问题，随着电子技术的发展，我们还可以采用计算机动画技术，对颜色混合实验进行计算机模拟。

3. CIE 标准色度系统

根据三基色原理，光谱上的所有颜色都可以由红(R)、绿(G)、蓝(B)三种颜色混合产生。色度是用红、绿、蓝三基色匹配产生可见光谱中不同颜色时的三基色比例系数。国际照明委员会(简称 CIE)据此制定了一个色度规范的标准测量系统，称为 CIE 标准色度系统。任何一种颜色都可以用红、绿、蓝三基色按特定比例混合得到。假如三基色的比例以光度单位表示，可以得到颜色方程式：

$$(\mathrm{C}) = r(\mathrm{R}) + g(\mathrm{G}) + b(\mathrm{B})$$

其中，C 指某种特定的颜色，R、G、B 表示三基色，r、g、b 分别表示每种基色的系数，“＝”表示匹配，即视觉上颜色相同。

因为三基色的总光量必须与被表示的颜色相等，所以 $r+g+b=1$。如某种蓝绿色可表达为：

$$(\mathrm{C}) = 0.06(\mathrm{R}) + 0.63(\mathrm{G}) + 0.31(\mathrm{B})$$

用三基色匹配某一颜色时，r、g、b 中的有些值可能会取负值，这等于是说要在被匹配的颜色上加上某种颜色，才能达到相等的颜色匹配。如要得到橙色的等色，其方程式可表达为 $(\mathrm{C}) = r(\mathrm{R}) + g(\mathrm{G}) - b(\mathrm{B})$，这等于是说将 R 和 G 按照 r 和 g 的分量混合时得到的橙色与目标橙色不匹配，要想达到匹配，需要在目标橙色中加上 b 分量的 B。

1931 年，国际照明委员会根据上述原理，用 X、Y、Z 分别代表红色、绿色和蓝色，而用 x、y 和 z 代表这三种颜色各自所占的分量。由于 $x+y+z=1$，当 x、y 确定后，z 分量也就确定了。使用红光和绿光的分量 x、y 可以确定一个颜色的二维坐标位置。经过对各种颜色分量的大量测定，国际照明委员会制定了一个色度分析系统，又称 CIE 色度图或 CIE1931 色度图，如图 6-5 所示。在色度图上，只要找到一个颜色的坐标位置，就可了解其组成的三基色分量，还可看到

此颜色的补色、明度和饱和度。CIE 色度图的横坐标为 x,代表红光分量;纵坐标为 y,代表绿光分量。图中没有 z 色度坐标,但它可由公式 $z=1-(x+y)$ 推算。光谱的红色波段在图的右下部,绿色波段在图的左上部,蓝色波段在图的左下部。图中曲线表示光谱颜色的轨迹点,代表从 380～780 纳米的纯光谱色,图中曲线以内的任何一点则不代表纯光谱色,而是几种光谱色混合得到的某种颜色。这个色三角形内包括了一切物理上能实现的颜色,落在色三角形以外(即光谱轨迹以外)的点都是不能实现的颜色。

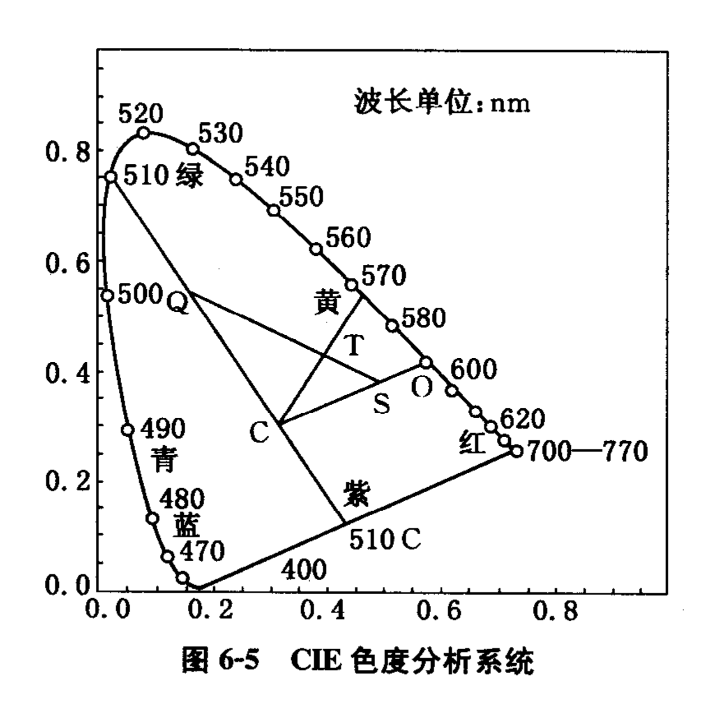


图 6-5 CIE 色度分析系统

图中 C 点是一个指示点,近似昼光的标准白光,它包含的三基色的比例几乎相等,其中 $x=0.3101$, $y=0.3163$,从而得到 $z=0.3736$。任何一种颜色都可以在色度图中用它所占的位置来表示它的特性。图中的 S、Q 表示两种颜色,其坐标分别是 S(0.48, 0.37)、Q(0.16, 0.55)。从 C 点到光谱轨迹上可以任意作许多直线,在同一条直线上的各点所代表的色光均属于同一主波长,①并代表着不同程度的饱和度。因为 C 点代表白光,越靠近 C 点的颜色,其饱和度越小,反之,越接近光谱轨迹的颜色,其饱和度就越大。

从色度图还可以推算出由两种颜色混合得到的各种中间色。如颜色 S 和 Q 相混合,可以得出 S 到 Q 直线的各种颜色,以此直线上的 T 点为例,可以从 C 点作一直线通过 T 点交于光谱轨迹上的 572 纳米处,则可由 572 纳米看出 T 点的

① 一个物体所反射的最大能量的波长部位就是该物体的主波长。

色调是一种黄绿色,并可从T点到C点的距离看出T点颜色的纯度。

二、暗适应的实验研究及其应用

感觉适应是指感受器在一定强度的刺激物持续作用下感受性发生变化的过程,就视觉来说,是指明适应或暗适应。当人们从光亮环境进入较黑暗环境时,刚开始会感到周围漆黑一片,什么也看不到,但只要周围还有一些微弱的光存在,等待片刻,人们就能逐渐地看到些周围物体,能见度不断提高。在暗视条件下,视觉反应的阈值下降,光敏感度逐步提高的现象称为暗适应。暗适应过程就是由于光刺激由强到弱的变化所引起的视觉分析器感受性提高的过程。

(一) 暗适应过程

暗适应(dark adaptation)主要是感光细胞的暗适应,包括视锥细胞和视杆细胞暗适应,其中前者适应过程较快,但感受性提高的幅度有限,最大的提高幅度大约能达到适应前的几十倍;后者适应过程较慢,但感受性提高的幅度很大,最大的提高幅度大约能达到适应前的几十万倍以上。由于暗适应过程中,视觉功能从中央视觉转移到边缘视觉,即从中央窝的视锥细胞转移到中央窝以外区域的视杆细胞,而且视杆细胞的适应过程长、适应幅度大,因此暗适应的主要机制在于视杆细胞的感受性逐步提高。

当人进入黑暗环境后,很快就会发生视觉绝对阈限降低,眼睛的感受性增加,但是其增幅不大,这是视锥细胞的暗适应,视锥细胞的暗适应大约只需要5～7分钟即可基本完成。在这段时间内,视锥细胞的感光色素虽然在加速合成,光敏感度有所提高,能感受到一些微弱的光亮,但暂时还看不清光度很微弱的物体。随着时间的推移,视杆细胞中的视紫红质(visual purple)含量逐渐恢复到足以使视杆细胞兴奋的水平,使暗视觉的光敏感度达到应有的水平,这是视杆细胞的暗适应过程,大约需要20～30分钟后才能完成。虽然视杆细胞的暗适应时间长,但是适应程度很高。

在典型的暗适应实验中,一般要使被试先达到高亮度的明适应状态,方法是要求他用几分钟时间去注视一个数百英尺朗伯亮度的均匀的白色屏幕中心。然后,再用不同亮度的测试光来测定其不同时刻的明度感觉阈限,即在黑暗环境中测定其视觉绝对阈限与时间的函数关系,考察其眼睛的暗适应进程,以及这一过程中的规律性。

(二) 暗适应曲线

上述暗适应过程可以表示成一条二维曲线:横坐标表示进入黑暗环境的时间、纵坐标表示眼睛绝对感觉阈限的相对水平,由此形成的二维曲线就叫做暗适应曲线,如图 6-6 所示。暗适应曲线反映的是人在进入黑暗环境后,随着时间的推移,其眼睛绝对感觉阈限逐步下降的过程,也就反映了其眼睛感受性逐步提高的过程。如果用白光测定眼睛的暗适应过程,就会得到如图 6-6 所示的暗适应曲线,这条暗适应曲线由两个部分组成,正反映了上述的两种暗适应机制:最初的一部分曲线下降很快、下降幅度较小、在 7 分钟左右即趋于稳定,反映的主要是视锥细胞的适应过程;后一部分下降幅度很大,需要的时间也较长,在 30 分钟后才逐渐趋于平缓,这反映的主要是视杆细胞的适应过程。如果用红光照射中央窝并检测其绝对感觉阈限,结果发现只能得到图 6-6 中上部的较为平坦部分的曲线,即视锥细胞的暗适应曲线,但得不到视杆细胞的暗适应曲线,因为这种细胞对红光的敏感性较差,几乎不会对其发生适应性反应。

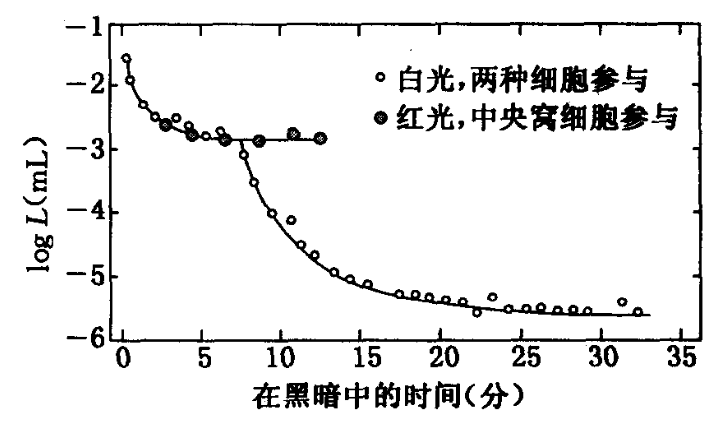


图 6-6 人眼的暗适应过程和曲线

注:纵轴是达到视觉绝对阈限时测试靶的亮度的对数;
横轴是眼睛离开高强度光视野后经历的时间。

(三) 暗适应的影响因素及应用研究

1. 暗适应过程的影响因素

人的暗适应过程受到多种环境因素和机体因素的影响和制约,概括地说,主要包括三个方面:前适应状态、检测光的刺激模式和有机体的自身状态。

前适应状态就是指眼睛在进入黑暗环境之前接受的光照强度和光照时间,它使得眼睛处于某种适应状态。一般是光照强度越大、时间越长,眼睛的感受性越低,进入黑暗环境后的暗适应过程就越长。据沃尔德(Wald, 1944)的研究,认

为这种现象与视紫红质的光化学反应有关:①在视杆细胞中存在一种感光物质,叫做视紫红质,它在曝光时很快被破坏退色,在暗适应过程中又重新合成并恢复活力。视紫红质发生光化学反应的方程式是:

视紫红质=视黄醛+蛋白质

这是一个可逆的化学反应过程。在光刺激时,视紫红质发生分解而退色,变为黄质(视黄醛+蛋白质)。当光继续作用时,视黄质再进一步分解而退色,变成视白质(维生素A+蛋白质)。就是说,眼睛受到光刺激的时间越长,视紫红质分解的就越彻底。相反,暗适应时视紫红质重新合成的时间越长,完成暗适应的过程也就越长。

在暗适应过程中,检测的光刺激模式不同,暗适应过程也不同,这主要包括检测光的颜色、刺激视网膜的位置和面积大小等。比如,检测暗适应过程中使用的是红光,或者刺激的位置在中央窝的很小区域,那么发生适应的就只有中央窝的视锥细胞,其过程时间短,但变化幅度小;如果检测暗适应的是白光且作用于视网膜的边缘区,那么检测到的就是视杆细胞的适应过程;如果检测暗适应的刺激是白光,且作用于包括中央窝的较广大区域,那么两种视觉细胞都会发生适应,就会出现像图6-6所示的适应过程。一般而言,光刺激面积越大,暗适应速度越快;刺激视网膜中央窝及其临近区域,暗适应过程就有一定的限度,而越靠近视网膜边缘,暗适应速度就越快。有研究发现,在同等亮度下绿色光对暗适应的影响大于红色光。绿色光刺激强度越大,暗适应时间越长。光刺激作用主要集中在视网膜中央窝的条件下,等亮度较弱的红光与绿光之间暗适应时间无显著差异。②

最后,暗适应过程还与机体状态有关,其中最重要的是与疲劳程度关系密切。在疲劳程度较高的条件下,暗适应的水平会降低,因此暗适应水平可以作为个体疲劳程度的一个测量指标。暗适应的个体差异也非常明显,这又与其以往的生活习惯和经历有关,一个从小就经常生活在较暗环境中的人,其暗适应性就比较强。根据视杆细胞的光化学反应机制,暗适应也与个体维生素A是否缺乏有关。当比较缺乏维生素A时,视紫红质合成的速度和总量就比较低,暗适应就显得较为困难。

---

① 转引自:张庆云:《普通实验心理学》,河南大学出版社1993年版,第174页。

② 许百华、傅小贞:《液晶显示的红、绿色光对视觉暗适应影响的比较研究》,《人类工效学》,2002,8(4):17—19。

2. 暗适应现象及其应用研究

人在暗适应条件下,有时会出现一些有趣的现象,当然有时也会出现一些对人的安全、完成作业等不利的情形,这就需要对之开展一些应用性研究,以便在查明暗适应现象发生机制的同时,预防其中的不利因素。

在暗适应状态中,有两种现象比较典型:"斜视"现象和"浦肯野树"现象。当我们在黑暗中停留一两个小时后,直接看一个很微弱的测试光点或星星时,反而会看不到了,这是由完全的暗视觉后效造成的,因为中央窝没有视杆细胞,视锥细胞此时几乎不能发生作用,所以在完全暗适应的条件下中央窝便成了盲区。如果这时不直接看微弱光点,而是斜视,即用中央窝以外的区域去看时,就能观察到微光下的物体。学会用中央窝以外的区域去看东西是观测天文学家和夜间外出执行任务的人必须掌握的一种技术。

另外,在暗适应条件下,还可看到一种"浦肯野树"现象。这种现象就是你能看到你自己的视网膜上的血管,当把一个点光源放在你眼前时,这些血管的影子便或多或少地落在视杆细胞或视锥细胞上。当你用一个小光源从侧面照射你的眼睛时,就能很清楚地看到这种现象。沿着一条小弧线在眼睛的一侧很快地移动小光源,就会看到"树"的图像。树的枝和杈代表着玻璃体与视网膜之间粗细不同的血管,而"树干"就在盲点的位置上,这也是血管进入眼球的地方,同时也可以看到中央窝区域没有血管。尽管视网膜上血管很多,但在平时的视觉中,血管的影子是看不到的。

与暗适应有关的应用研究很多,其中比较典型的有三个方面:疲劳测试、眩光克服和红目镜的使用问题。首先是疲劳测试问题,这一点在前边已有所讨论。人在疲劳情况下,暗适应的速度会降低,能达到的暗适应水平也会降低,因此可以反过来把暗适应状况作为疲劳程度的度量指标,尤其在视觉实验过程中,它能很好地反映眼睛的疲劳程度。

其次,关于眩光(glare)的克服问题。眩光是由光源或反射面的亮度过大,或由光源或反射面与背景的亮度反差太大造成的视觉现象,可引起视觉的不舒服感,降低视觉工效。[①]引起眩光的光源或反射面叫做眩光源。根据眩光源的不同,可分为直接眩光和反射眩光;根据对人的视觉影响程度,可分为不舒服眩光、失能眩光和失明眩光。在设计和改善照明环境时,应尽可能地控制眩光的产生。

---

① 林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社 2003 年版,第 1478 页。

对于汽车驾驶员来说,夜间行车中,突然遇到强光照射会出现暂时性的视力下降,给行车安全带来危害,所以也要尽量避免这种强眩光的出现。比如,公路设计中,中间以低矮植物形成隔离带,可以在夜间行车中有效地遮挡对面逆向行驶汽车车灯的影响;夜间行驶中,尽量不要以大灯照射对面车辆或行人,以免形成眩光等。除交通领域,城市广场照明设计,甚至学生使用的台灯都要考虑到这一点。不过,抓老鼠时倒可以利用眩光,如以手电筒直接照射老鼠,就会使得老鼠出现眩光反应,"静坐"待毙。

最后,利用感光细胞的暗适应特点和规律,保证夜间执行任务者的有效照明——使用红光照明或佩戴红目镜。红目镜可以透过波长长于620纳米的红光,而滤掉波长短于620纳米的光线,它在光线较强的环境下,既可以减弱光线的强度,又只能使红光才能进入眼睛。图6-7中阴影部分表示的区域为红滤光片允许光通过的区域,表明红光仍可以相当有效地刺激视锥细胞,这样,人们仍有良好的视觉能力。黑色部分表示的区域为红光刺激视杆细胞的区域,该区域很小,说明红光几乎不能刺激视杆细胞。因此,戴上红目镜后,使得视杆细胞如同进入暗适应状态一般。对军警夜间执行任务、工业生产、医务检测、照相暗室等经常需要在明暗处交替的工作而言,必须戴上红色的眼镜,或者用红光照明,从而保证视杆细胞相对地不受光线的强弱变化而保持良好的暗适应状态,以减少暗适应时间,提高工作效率,增强人身安全。

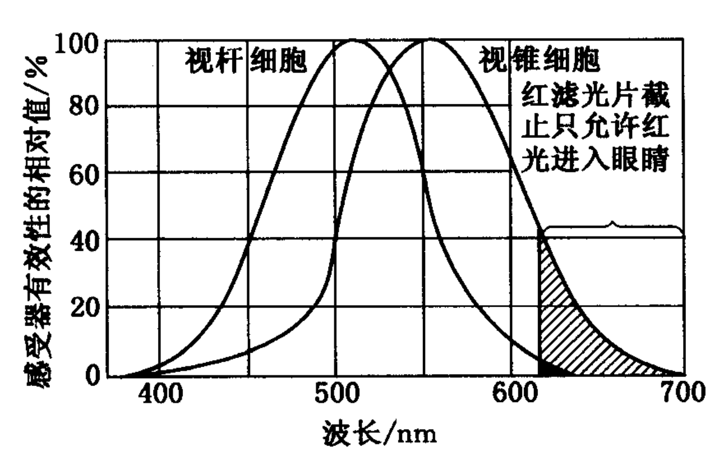


图6-7 红目镜滤光作用示意图

三、视敏度的测量及影响因素

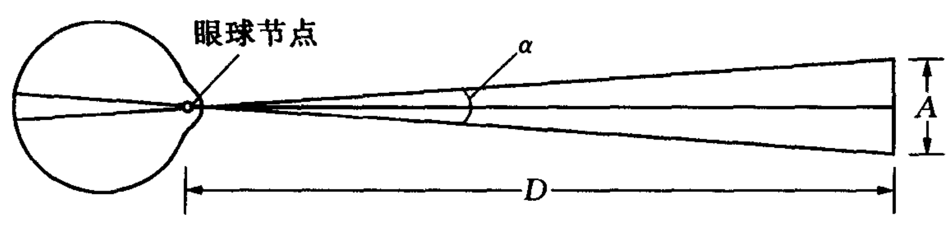
视敏度(visual acuity)是眼睛的分辨能力,包括空间视敏度和时间视敏度两

种。空间视敏度是眼睛对刺激物空间细节大小的分辨能力，传统的视力测试都是测量空间视敏度；时间视敏度是眼睛对时间间距的分辨能力，通常用闪光融合临界频率来测量。视敏度测量不仅是心理学研究的基本课题，而且是人们在社会和日常生活中经常遇到的活动。这里对视敏度的测量方法、视敏度的影响因素进行分析。

（一）空间视敏度

空间视敏度，体现了视觉的空间辨别特性，一般简称为视敏度或视力。空间视敏度是测量人对空间刺激物的分辨能力，是指人眼分辨物体细节的最大能力，常以能看出来的物体细节的间距来表示。一个人辨别物体细节的间距越小，视敏度就越高，反之越低。

空间视敏度通常用能分辨的最小空间细节形成的视角来确定，视角就是物体大小对眼球光心形成的夹角，其大小决定了刺激物在视网膜上的成像大小。一般情况下，正常眼能分辨的最小空间距离形成的视角约为 1 分(即 1/60 度)。很明显，眼睛能分辨的空间细节越小，其形成的视角也越小，反映其眼睛的分辨能力就越强，所以眼睛的空间视敏度与这一视角的大小成反比。如果用 V 表示眼睛的空间视敏度，用 α 表示能分辨的最小细节形成的视角(单位：分)，则 $V=1/\alpha$。视角的测量方法如图 6-8 所示。实际应用中，一般使用“兰道环(Landolt ring)C”或 E 形视标作为测量工具，即在一定光照条件下，要求被测者站在离开视标一定距离处(标准观察距离是 6 米)，用单一裸眼或矫正眼观察视标，并报告其看到的视标的开口方向。不断变换视标开口的方向和视标大小，找到观察者能正确识别的最小视标，这一视标开口的大小就是被测者能分辨的最小空间距离，相当于图 6-8 中的 A。得到 A 后，利用视敏度计算公式就可得到该被测者的视敏度。



其中：α 为视角　A 为眼睛分辨的最小空间距离

D 为视标与眼睛节点的距离

$\alpha = A/D$(弧度) $= 57.3A/D$(度) $= 3\,438A/D$(分)

视敏度 $V = 1/\alpha = D/(3\,438A)$

图 6-8　空间视敏度测量方法示意图

空间视敏度既是一种生理指标,又是一种心理指标,它受到多种生理因素、心理因素和刺激因素的影响和制约,具体地说,主要有以下几个方面:

(1) 光照条件。在不同的光照环境中,眼睛的视敏度会有所不同,比如在不同亮度水平上,视锥细胞的适应水平不同,其分辨物体细节的能力也不同;不同色光照明条件下,眼睛的适应水平或分辨能力也都不同,视敏度也就有所不同。有研究表明,随着刺激物亮度增加,空间视敏度也增加,二者呈对数关系,但当视敏度达到 2.0 以后便趋于稳定,不再随亮度的增加而提高。

此外,刺激物与背景亮度的对比关系也影响视敏度。当物体与背景之间亮度的对比度加大时,空间视敏度提高;反之,则降低。特别是当背景光过于强烈,造成观察者的眩光,其视敏度就会降到非常低的水平。当然,刺激物本身的亮度过大,也会形成眩光,造成视敏度下降。

(2) 视网膜部位。视网膜的不同区域具有不同的视敏度。在中央窝处,视锥细胞密集并且比较纤细,又与双极细胞和神经节细胞形成单线联系,因而对物体形态的精细分辨能力最强,因此中央窝处的视敏度最大。越靠近视网膜边缘区,视敏度越低,因为视网膜边缘部位主要是视杆细胞,视锥细胞较少且分散,视杆细胞与双极细胞和神经节细胞形成汇聚式联系,在视觉信息传递中进行信息整合,这种整合虽可以对某些信息进行放大,但也会造成某些信息的丢失,比如造成物体空间细节信息的丢失,因而视敏度较低。

(3) 观察者的身心状况。人在良好的身心状态下,视敏度会比较高;在疲劳(包括生理上的疲劳和心理上的疲劳)、有生理疾病、缺乏营养(比如缺乏维生素 A)等,都会造成视敏度降低。当然,一定的练习和良好的视觉适应,也可以促进视敏度水平的提高。

(4) 观察者的年龄。在青少年成长过程中,随着年龄的增加,视敏度水平逐渐提高,成年后处于稳定状态,40 岁以后,特别是 50 岁以后,视敏度开始有所下降。一般来说,人在 14～20 岁时的视敏度最高、20～40 岁时的视敏度比较稳定、40～65 岁时的空间视敏度只相当于 14～20 岁时的 1/4～1/3,60 岁以后迅速下降。当人逐渐衰老时,晶状体逐渐变硬,失去弹性,而使眼睛的自动调节能力降低,以致晶状体表面不能形成应有的曲率,使近距离物体的视像不能很好地聚焦在视网膜上,所以随着年龄的增加,视觉的近点逐渐变远,瞳孔面积也随年龄增加而缩小。一般来说,进入老年阶段后,年龄越大,视敏度越低。

(二) 时间视敏度

时间视敏度,是眼睛分辨时间间距的能力。一般的测量方法就是要求被测者观察两个先后呈现的视觉事件,找到被测者能够分辨的最小时间间距。在心理学实验室里,可以使用闪光融合临界频率作为眼睛时间视敏度的测量参数。

如果一个间歇频率较低的光刺激作用于人眼时,会产生一亮一暗的闪烁感,但随着闪光刺激的频率逐渐增加,其先后两次闪烁的时间间距越来越小时,人们会逐渐地将这一闪烁光点看成是稳定的光点,这种现象叫做闪光融合。例如,我们照明用的日光灯每秒钟闪动 100 次,但我们并不觉得灯光是闪动的,这实际上是一种闪烁光的融合状态。产生闪光融合时的闪光频率称为闪光融合临界频率(critical fusion frequency,简称 CFF)。

光刺激人眼后,视网膜的反应在时间上有一迟滞,从而产生视觉后象。闪光融合现象的产生与视觉后象有关。一般来说,在中等强度的情况下,视觉刺激后象保留的时间大约为 0.1 秒。这种时滞的存在对于我们知觉物体是一种优点。要使不连续的刺激引起连续的感觉,就必须使它达到一定的频率。一般而言,闪光频率达到 30～35 Hz,闪光可能会被知觉为连续的,就不再有闪烁感。当其他条件相同时,闪光融合临界频率越高,就表明时间视敏度越好,即眼睛对于时间上的明暗变化的分析能力就越强。

时间视敏度的测量本身还缺乏非常有效的方法,虽然 CFF 可以在一定程度上反映时间视敏度的水平,但由于其测量本身稳定性比较低,所以对时间视敏度作精确的研究尚存在困难。关于时间视敏度的生理机制,还缺乏足够的研究资料,还不能得到被广泛接受的解释。该领域研究比较多的,还是关于其影响因素的研究,尤其是闪光融合临界频率与闪烁光点的强度、颜色等的关系的研究。关于时间视敏度影响因素的研究,我们介绍以下几个方面:

(1) 刺激的光相强度。闪光在时间和强度上可分为暗相和光相。假定暗相光强为零,则闪光融合临界频率与光相强度的对数成正比,其数学公式为:

$$CFF = a\lg I + b$$

其中,CFF 为闪光融合临界频率,I 为光相强度,a、b 为常数。这一关系是费里(Ferry, E. S.)和波特(Porter, F. C.)研究发现的,因此被称为费里-波特定律。不过,这一定律只在中等强度的光相时才成立。一般而言,随着光相强度的增加,闪光融合临界频率也增加。但是,在当闪光融合临界频率低于 5 Hz,或高于

50～55 Hz时，此公式也不能成立。

(2) 刺激光点的颜色。刺激光点的颜色不同，*CFF* 也有所不同，一般来说，红光的 *CFF* 较高。

(3) 刺激的时间特性。这主要是指闪光中光相和暗相的时间比，光相时间相对越长，光相的后象效应就越大，且两次闪烁间的暗相时距就越小，此种条件下测得的闪光融合临界频率会比较小；反之，如果光相时间相对较短，则测得的闪光融合临界频率较高。

(4) 刺激的空间特性。这主要是指光刺激的面积。由于刺激的空间累积作用，面积较大时的闪光融合临界频率比面积较小时的要高些。闪光融合临界频率随着闪光照射的面积的扩大而增大，二者也存在着对数关系，数学公式为：

$$CFF = c\lg A + d$$

其中，*CFF* 为闪光融合临界频率，*A* 为光刺激的面积，*c*、*d* 为参数。

(5) 刺激的网膜位置。由于视杆细胞和视锥细胞的分辨能力有很大差异，所以当闪烁光刺激的视觉细胞不同时，其闪光融合临界频率也不同。当刺激中央窝区域时，闪光融合临界频率比刺激网膜的边缘区时要高。

此外，有机体的身心状态如年龄、疲劳、缺氧、药物和酒精等作用，以及双眼间的迁移等因素都会影响闪光融合临界频率。

最后，需要指出的是，眼睛的空间视敏度和时间视敏度具有一定的内在联系。通常，不论是优势眼还是非优势眼，人眼的时间视敏度和空间视敏度都有着极其显著的正相关。通过有针对性的训练，可有效提高个体的空间视敏度，从而提高个体的整体视力水平。[①]要想充分了解二者的关系，必须综合考察其中的各种影响因素以及各因素间的相互作用，才能对视敏度的实质有一个正确的理解，并有助于建立一个全面、客观、科学的视敏度测评体系。在目前的实验心理学研究领域，对于这两种视敏度的研究往往仅限于就某一种特别的空间视敏度的影响因素做单一讨论，尚较少涉及这两种视敏度的相关关系的研究。

第三节　耳的结构与听觉测量

人类靠耳获得的信息虽远不及靠眼睛获得的信息量大，但耳也是我们获取

---

① 周海谦：《空间视敏度与时间视敏度的相关关系研究》，《许昌学院学报》，2004，23(2)：57—59。

信息、了解世界的最重要通道之一。就心理学研究而言，由于多方面原因，特别是研究技术方面的原因，在听觉方面积累的文献相对少一些，有关感觉和知觉的许多知识都来自视觉研究。作为知识基础的铺垫，这里对听觉的适宜刺激、耳的结构功能、听觉的测量与分析进行概括性介绍。

一、听觉的适宜刺激

听觉(audition)的适宜刺激是声波(sound wave)，它是物体振动在弹性介质中激起的一种纵波(vertical wave)。听觉就是由物体振动产生的声波作用于人(或动物)的听觉器官——耳朵后产生的一种感觉现象。当我们敲击一个音叉时，它会连续地前后振动，使周围空气分子发生位移，形成空气分子的压缩(正压)或扩散(负压)的压力波。正、负压在空气这种弹性物质中形成交替运动并向四周传播，当空气振动的波进入内耳(inner ear)并引起内耳鼓膜振动时，听觉系统会将振动的机械能转换为神经能并传达到听觉中枢，这时就产生听觉。

声波的物理特征包括频率(frequency)、振幅(swing)和波形(wave shape)，其中频率反映了声源的振动快慢，振幅反映了振动波的能量，波形则反映了波的成分。一般来说，人耳能接收并产生声音感觉的声波频率范围为 16～20 000 Hz。[①]当声波的频率超过 20 000 Hz 时，即使它作用于人耳也不能形成声音感觉，所以声学上把它叫做超声波(ultrasonic waves)；当声波的频率在 16 Hz 以下时，即使它作用于人耳，但大部分人也不能形成声音感觉，所以声学上将这样的声波称为次声波(infrasonic waves)。日常工作和生活中，人们每天都要接收大量声波刺激，根据声波的波形特征和听觉效应，可以将听觉刺激划分为两大类：纯音(pure tone)和复合音(complex tone)。复合音又包括乐音(musical tone)、噪声(noise)和语音(voice)三种类型。

纯音是单一频率的声波，波形为正弦波。在心理学实验室中，可使用音叉振动来产生这种纯音，如图 6-9 所示的就是由音叉产生的声波波形。振动的音叉导致周围空气分子的振动，它们振动的幅度可以用一个正弦波表示。周围空气中任何一点从一个时间到另一个时间压力的振动也是一个正弦波函数。同样，

① 也有很多文献介绍说声波的可听范围是 20～20 000 Hz，但是心理学研究发现，在比较精细的研究中，不少被试的音高绝对阈限可以达到大约 16 Hz。

这种振动在空气中传递到内耳时,推动我们鼓膜前后运动的空气压力变化导致鼓膜也以一种简单正弦曲线函数振动。在鼓膜那里,压力大小的变化明显地与音叉臂往复运动的幅度有对应关系。不过,在实际生活中,我们能听到的声音极少是单一的纯音,绝大多数都是由不同频率和振幅的声波混合而成的复合音。

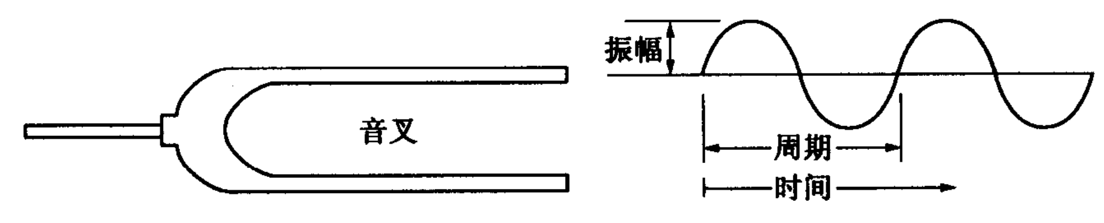


图 6-9　音叉及其振动产生的纯音波形

复合音是由不同频率的多种声波复合而成的复杂声波。任何复合音都可以分解为两种或两种以上的频率不同的纯音,按照组成复合音的各纯音的频率之间的关系,可以把复合音进一步划分为三种类型:乐音、噪声和语音。

乐音,"波形呈周期性变化的声音,通常指听起来和谐悦耳的纯音和由具有谐波关系的纯音所组成的复合音"。① 所谓的谐波关系,就是指组成乐音的各纯音的频率之间具有简单的倍数关系,即其中有一种频率最小、振幅最大的基音和一种以上的频率为基音频率整数倍的陪音,这些纯音复合在一起,依然可以构成一种有周期性变化规律的复合音。

噪声也由多种纯音复合而成,但由于这些纯音之间没有谐波关系,所以组成的复合音不再具有周期性,呈现出不规则的波形。人耳在接收到噪声作用后,由于其振动的不规则性,也会在一定程度上引起相应神经系统的活动紊乱,产生不良作用,所以噪声常常引起人们的情绪烦躁。噪声超过一定强度(85 分贝),较长时间作用于听觉器官,就会影响人们的工作效率和健康。当然,噪声也有它的应用价值,如临床上常采用白噪声(white noise)来测量耳朵的听力。白噪声是一种特殊的噪声,它是由各种振幅接近、频率不同的声波复合而成的噪声,其中的各种声波成分不具有谐波关系,在相位上也无系统关系。之所以称之为"白噪声",是从白光类比而来的:当各种频率的光波按照大致相等的强度混合在一起就可以形成白光,于是就把各种频率不等、振幅相等或相近的声波复合而成的噪声叫做"白噪声"。

① 林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社 2003 年版,第 1617 页。

语音是乐音与噪声的复合音,由元音(vowel)和辅音(consonant)两部分构成,其中元音是一种能连续发出的乐音,具有周期性;辅音主要是不能连续发出的短促的噪声,不具有周期性。元音与辅音复合组成语音音节,波形具有一定的规则性,但不是完全的周期性,图 6-10 就是元音"a"的一小段波形。

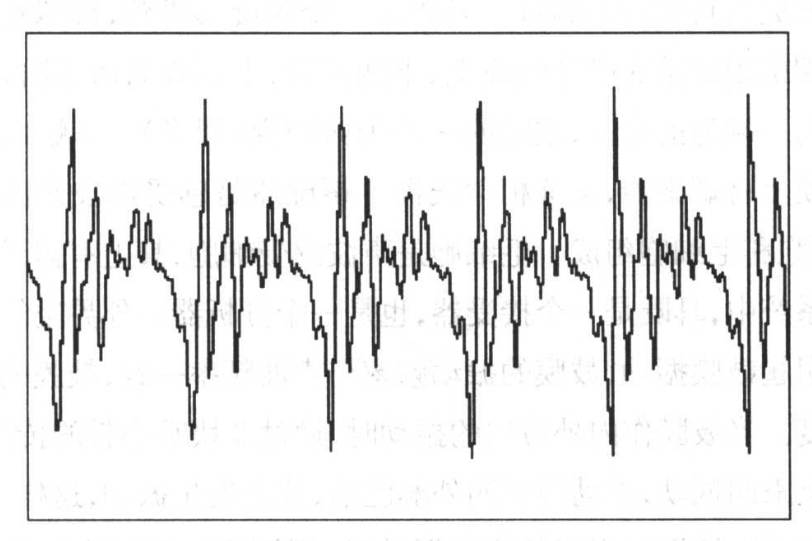


图 6-10 元音"a"的波形图

二、听觉信息的接收、传递与测量

(一) 耳的结构与机能

听觉的感觉器官为耳,耳由外耳、中耳和内耳三部分组成,其结构如图 6-11 所示。

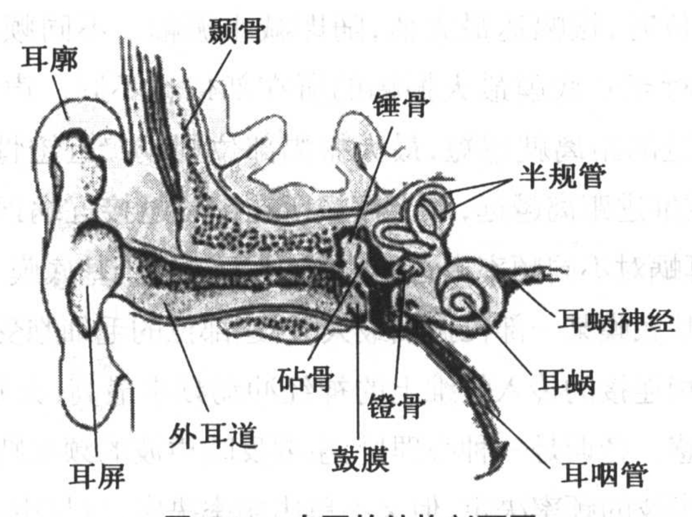


图 6-11 人耳的结构剖面图

外耳包括耳廓、外耳道和鼓膜。耳廓,呈漏斗形,有助于收集外界的声波,以

辨别声音的来源。经外耳道传来的声波,引起鼓膜发生振动。

中耳包括鼓室和耳咽管。鼓室内有3块听小骨(锤骨、砧骨和镫骨)。耳咽管通过闭合可以使鼓室内气压和外界气压相等,以保证鼓膜的正常振动而维持正常听力。

内耳由前庭器官和耳蜗组成。耳蜗是一个螺旋形骨管,形似蜗牛壳。骨管内部被骨质螺旋板和基底膜分隔成上、下两部分,上部称为前庭阶,下部称为鼓阶。前庭阶的一端为前庭窗,鼓阶的一端为卵圆窗(圆窗)。耳蜗内部充满着淋巴液。基底膜上有螺旋器,又称柯蒂氏器。螺旋器是感受声波刺激的听觉感受器,由支持细胞和毛细胞组成。毛细胞是声波感受细胞,其上有盖膜。

在听觉系统中,耳既是一个接受器,也是一个分析器。外界空气中的声波经外耳道传入引起鼓膜振动,鼓膜的振动频率与声波频率一致,鼓膜的振幅也决定于声波的强度。当鼓膜作内外方向的振动时,通过3块听小骨的传递,使镫骨底板在前庭窗上来回振动,推动内耳的外淋巴液,使之发生振动,这样,声波就传进内耳。当内耳淋巴液振动引起基底膜振动时,螺旋器的毛细胞与盖膜的相对位置不断发生变化,可能因此引起毛细胞发生膜电位变化。毛细胞的膜电位变化进一步又引起耳蜗神经纤维产生动作电位。神经冲动传至大脑皮质听区而引起听觉。

那么,声波的特征如何在感觉中得到反映呢?现代生理学的研究发现,当声波通过镫骨底板作用于前庭窗使耳蜗淋巴液振动时,基底膜是以一种行波方式运行的。行波从蜗底开始,向蜗顶方向推进,振幅也逐渐增大。当行波到达基底膜某一部位时,振幅达最大值,随即减小振幅。不同频率的声波,从蜗底起始的行波运行至基底膜最大振幅的所在部位也不同。声波频率越高,行波在基底膜上推进的距离就越短,最大振幅部位则越接近镫骨底板。反之,声波频率越低,行波推进距离越远,最大振幅所在部位越接近蜗顶。因此,听觉的行波理论认为,耳蜗对不同频率声音的分析,主要决定于基底膜上行波最大振幅的所在位置。当基底膜某一部位振幅最大时,这部位的毛细胞受到的刺激最强,与此部位毛细胞相连接的传入纤维上的神经冲动频率最高,大脑皮质听区因此而获得某种音调感。音调是一种心理量,主要反映声波的频率特征,通常称为音高。音高主要由声波的频率决定,但又不单由频率决定,也与声波的振幅即强度有关。

当声波振幅增大时,虽然最大位移的基底膜部位不变,但是,位移的振幅也

增大,其结果不仅会使单根耳蜗神经纤维增加冲动频率,并且会使更多的耳蜗神经纤维向大脑发放冲动,有更多的神经元参与活动,从而在大脑皮质听区产生不同的响度感。响度感也是一种心理量,主要反映的是声波的振幅特征或强度特征,通常称为响度(loudness)。响度主要由声波的振幅决定,但又不单由振幅决定,也与声波的频率有关。

(二) 听觉的测量

听觉的适宜刺激是可听声波,而声波的特征与听觉的心理感觉直接相关,也就是说,听觉是对刺激声波特征的反映。如上所述,声波的主要物理量是频率和振幅,其中振幅决定了声波的能量强度,也因此决定了引起鼓膜振动的压力大小,我们称之为声压级。这样说来,在听觉测量中,主要是对频率、声压级两个物理量和音高、响度两个心理量的测量。

1. 听觉刺激的频率和声压级的测量

物理量的测量可以采用物理方法,如通过示波器等电子设备记录声波的波形,即可得到其频率和振幅参数。不过,在对听觉刺激的物理特征进行描述时,一般不直接使用振幅,而使用声压级,为了测量听觉刺激的声压级,需要规定其零点和单位。按照国际统一规定,不采用声波能量的实际测量值来表示声压级,而采用贝尔[①]或分贝(decibel,简称 dB)作为声压级的单位,同时给贝尔和分贝作出如下定义:“贝尔是所施加的功率或能量与某种参考的功率或能量的比率的对数。”[②]其中 1 贝尔等于 10 分贝,于是得到声压级的计算公式:

$$N(\mathrm{dB}) = 10\log(E/E_0)$$

公式中 N 代表声压级,单位为分贝;E 和 E_0 代表被测量的声波和参照声波的能量或功率。那么这个参照声波的能量如何得到呢?按照国际统一规定,将人们在中频音频段的平均绝对感觉阈限所对应的声压级作为 0 分贝,相应的声音强度(能量)E_0 对应的声压 $P_0 = 0.0002$ 达因/平方厘米。这样,在实际应用中,就可以使用声压计直接得到声波的分贝数。

目前,用于对听觉刺激的测量与分析的主要仪器包括:

---

① 使用“贝尔”作为声压级的测量单位,有纪念电话的发明者之意。亚历山大·格雷厄姆·贝尔(Alexander Graham Bell, 1847—1922),电话发明者,生于英国爱丁堡,后移居加拿大,再移居美国,对语音学有研究,并成功实现了利用电流传送声音信息,于 1876 年发明电话。

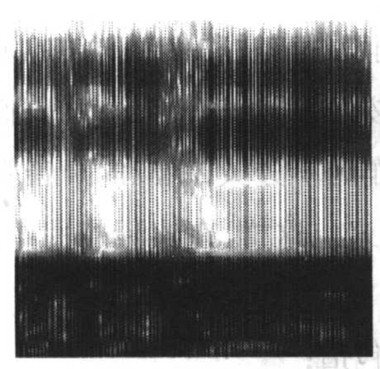
② 俞文钊:《实验心理学》,浙江教育出版社 1989 年版,第 253 页。

(1) 示波器(oscillograph)。用示波器可以记录和显示声波的频率和振幅。该设备能够通过换能器把声波转化为电波输入到加以调整的示波器后,可以显示为直角坐标系,横轴表示时间,纵轴表示振幅,得到的声波图形比较稳定而清晰。

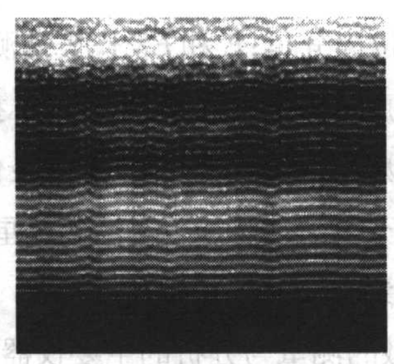
(2) 频率计(cymometer)。用频率计可以进行频谱测量。声波的频谱分析是将声波振幅的变化作为频率的函数来表现的。频率计可以将可听声的频率从 16～20 000 Hz 的高低相差 1 000 倍的变化范围划分成较小的频段。如果以频率为横坐标,以声压级为纵坐标,可以作直角坐标系进行声音的频谱分析。

(3) 声图仪(sonogram display apparatus)。用声图仪可以进行声图测量与分析。声图能从三个维度显示声波的振幅、频率与时间的函数关系,还可以标出不同频率成分的相位关系。这种测量方法和结果更为严密、精确,富有实体感。

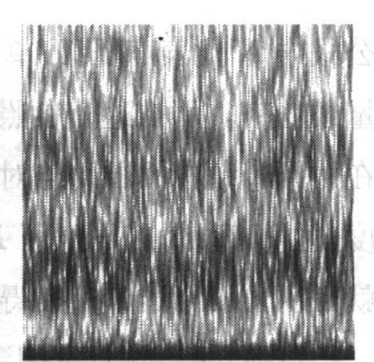
声图仪也叫语图仪,它能够将语音信息转换成一种可视图像,即把语音信息中的频谱成分及其能量分布表现在一个平面坐标系中,该坐标系的三维坐标分别是频率、强度和时间,其中横坐标表示时间,纵坐标表示频率,坐标系中散布的“灼烧”痕迹的浓淡表示强度或能量分布。图 6-12 所示的图像就是使用语图仪得到的元音“a”的宽带语图和窄带语图、白噪声的宽带语图。宽带语图能够比较清晰地看到语音材料中的能量分布,窄带语图则能清晰地看到语音材料中的音调变化,如图 6-12 中元音“a”的窄带语图显示该语音声调为“平声”。



元音“a”的宽带语图



元音“a”的窄带语图



白噪声的宽带语图

图 6-12　使用语图仪得到的宽带语图和窄带语图

南京师范大学心理学系建有专门的“语图分析实验室”(Sonogram Analysis Lab),装备了 3 套“VS99 语音工作站”,工作站装载的语音分析软件系北京阳辰

电子有限公司与中央司法鉴定中心联合开发的 Voice 3.0，可以实现语音编辑与分析完全软件化。该实验室可以进行各种语音实验材料的编辑与分析。

除上述软件与仪器外，还有声级计、声谱仪等，可用于对声音的波形、频率、音色等复杂成分进行分析和测量。随着电子技术的发展，已经有系统的电发声设备。如多种规格型号的信号发生器，包括高低频信号发生器、噪声信号发生器，同时还有各式录音机记录语言信号等。

2. 听觉的音高和响度测量

声波的特征量可以使用物理方法进行测量，但是音高和响度都是心理量，不能使用物理方法进行测量，而主要使用心理量表法。

音高也叫音调，主要由音频来决定。声波振动的频率越高，我们听到的音调就越高；反之，听到的音调就低，但它们之间是非线性关系，所以不能使用频率数直接度量音高。人们规定，1 000 Hz 的纯音在其声压级为 40 dB 时的音高为 1 000 mel，其他声波的音高则以此参考点进行测量，从而可以得到一个音高的主观量表。在这个主观量表中，音高与频率具有对数关系，音高量表实际上是一种对数式的等距量表，而且只有在 40 dB 的 1 000 Hz 刺激时，音高的 mel 数才正好等于其声波频率数。研究还发现，音高也受声音强度的影响。一般来说，"随着声强的增加，高频声(大于 3 000 Hz)显得更高；而低频声(小于 500 Hz)则显得更低……只有在中间频率时，音高才是声音频率的函数。而在低频或高频时，音高是声音频率和强度的函数。"①

响度是反映声音强度的主观量，但是声音响度不仅仅与声压级有关，而且与声音的频率有关。就是说，不同频率的声音，当具有相同的声压级时，它们的主观响度却不同。为了对响度主观量进行测量，需要对响度单位进行规定。不过，在文献和实际使用中，要注意区分两个概念，一个是响度，它反映声音有多响，其单位为 sone；另一个叫做响度级，它反映一个声音与另一个声音相比的响度等级，其单位为 phone。

(1) sone

sone 是国际上统一规定的主观响度单位，而且 1 sone 等于 1 000 Hz、40 dB 声音的主观响度，同时规定声压水平每增加 10 dB，则其主观响度增加到原来响度的 2 倍；声压水平每减少 10 dB，则其主观响度减少到原来的 1/2。如 1 000 Hz、

① 俞文钊：《实验心理学》，浙江教育出版社 1989 年版，第 272—273 页。

50 dB的声音的主观响度等于 2 sone；1 000 Hz、80 dB 的声音的主观响度等于 16 sone；1 000 Hz、20 dB 的声音的主观响度等于 0. 25 sone。按照这样的测量单位，声音的响度与其声压级呈指数函数关系，符合史蒂文斯定律（Stevens，1957）。

（2）phone

除 sone 外，还有另一主观响度单位——phone，这是常用在等响度曲线（后文详述）上的一个响度级单位，该单位规定：1 000 Hz 声音的分贝数就是该声音响度级的 phone 数。如 1 000 Hz、40 dB 的声音的主观响度级就是 40 phone；1 000 Hz、60 dB 的声音的主观响度级就是 60 phone。经大量测量后，在一个以频率为横坐标、声压级为纵坐标的二维坐标系中，可作出一系列的等响度曲线，如图 6-13 所示。在一条等响度曲线上的各个不同点，频率和声压级都是不相等的，但是它们的主观响度处在相同的响度级，这一响度级就等于该条曲线在频率为 1 000 Hz 的坐标点对应的纵坐标的分贝数（单位 phone）。如图 6-13 中的 A、B、C 点代表的声音的主观响度级分别为 20 phone、30 phone 和 4. 2 phone。其中 C 点处在零响度曲线上，其响度级是 4. 2 phone。

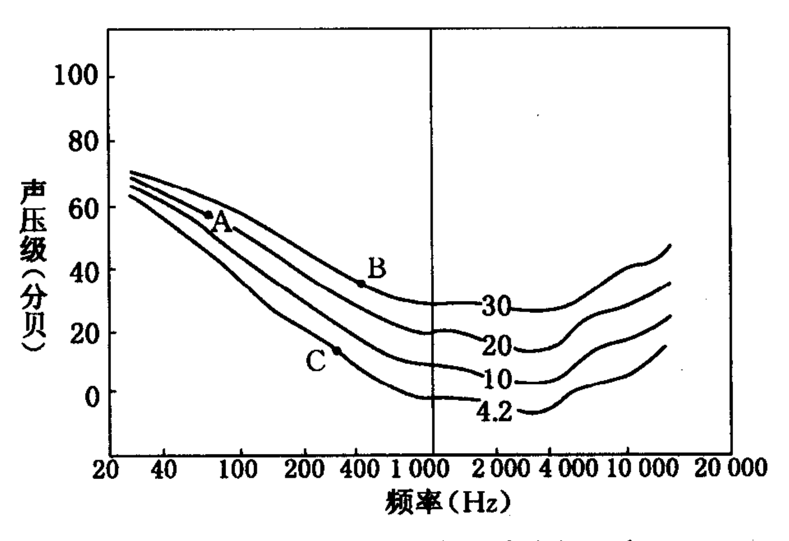


图 6-13　等响度曲线上主观响度级示意图

（A、B、C 点代表的声音的主观响度分别为 20、30 和 0 phone）

上述两个单位是按照不同方法规定的两种响度度量单位，这两种单位都具有相对性，均可用在反映声音的心理量方面，二者也具有一定的内在联系。比如 40 phone 的响度级正好是 1 sone 的主观响度、30 phone 的响度级也正是0. 5 sone 的主观响度。

第四节　听觉现象及其实验分析

一、音高测量及其与强度的关系

(一) 音高量表的制作

正如前文指出的，音高(pitch)主要与音频有关，所以音高量表主要是指音频与音高的对照表。音高量表常用的制作方法主要包括：两分法和等分法。两分法是让被试将一可变纯音的音高，调为标准音高的一半，并求出调整后纯音的频率，就可以得到音高与音频的一组对应值。等分法是给被试一个高频声 S_1 和一个低频声 S_5，让他在两者之间调出 3 个音 S_2、S_3、S_4，使临近两音，即 S_1 和 S_2、S_2 和 S_3、S_3 和 S_4、S_4 和 S_5 的音高的距离相等，如此得到一系列的音高与音频的对应值。使用这两种方法制作的音高量表非常一致。

在音高量表中，横坐标表示频率，纵坐标表示相应的音高，由此可以看出音高随声音频率变化而变化的函数关系，这种函数关系不是线性的关系，而是复杂的非线性关系。

(二) 音高与音强的关系

史蒂文斯等(Stevens et al., 1935)研究发现，音高不只与频率有关，也与音强有关。具体地说，当声音强度改变时，各频率声波的音高也会随之发生变化。对于低频音而言，音高随强度增加而降低；对于高频音而言，音高随强度增加而升高；对于某些中频音而言，强度的增减对音高只有轻微程度的影响，即音高基本上由音频决定。由于上述原因，关于音高的实验研究应该同声音的频率和强度结合起来进行，在说明一个声音的音高时，最好以一个标准的强度水平做参考，而这个常用的参照标准就是 40 分贝的强度水平。

二、响度测量及其与声频的关系

(一) 响度量表的制作

响度主要是声压级的函数，但也与频率和波形有关。史蒂文斯(Stevens, 1958)发展了一套直接测定响度的心理物理量表法。这种方法是让被试调整一

个 1 000 Hz 的纯音的强度,使它的响度听起来是 1 sone 的 2 倍那样响,这时的响度就定为2 sone。同理,也可让一个声音听起来是 1 sone 的一半,这时的响度即为0. 5 sone。依次类推,即可建立响度量表(又称 sone 量表),如图 6-14 所示。

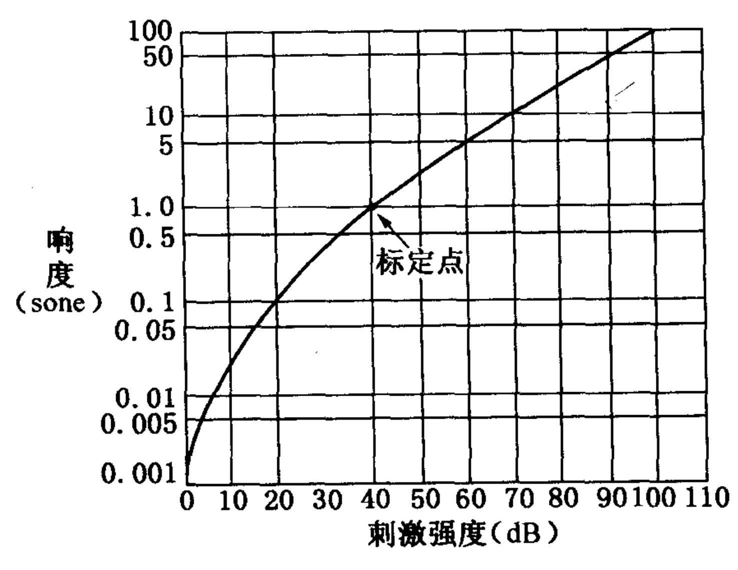


图 6-14 响度量表

(二) 等响曲线的测定

在以频率为横坐标、声压级为纵坐标的二维坐标系中,将响度级相等的各点连接起来即可得到等响度曲线。那么如何测定这样的等响曲线呢? 一般采用比较法:给被试先后两个声音刺激,其中一个是标准刺激(如 1 000 Hz、50 dB),一个是比较刺激(如 500 Hz、70 dB)。要求被试以标准刺激的响度为标准来调整比较刺激的声压级,直到被试感觉其与标准刺激一样响为止,记录调整后的刺激的频率和声压级,就可以得到一个坐标点。改变比较刺激的频率,仍以前边的标准刺激为准重新进行实验得到第二个坐标点,如此重复多次,就可以得到与标准刺激的响度相等的一系列坐标点,将这些点连接起来就得到一条等响度曲线。图 6-15 就是按照这样的方法得到的一系列等响度曲线,每一曲线处在不同的响度级上(10 phone、20 phone……)。从图中等响曲线的起伏可以看出,声音刺激的响度水平不单由声压级决定,也与其频率有关。在声压级较低的范围内,等响曲线呈"V"型起伏,说明响度与频率的关系也较为密切;在声压级较高的范围内,等响曲线趋于平缓,即在相同强度时有近似的响度,响度受频率的影响不明显。如果我们知道某一声音的频率和强度,就可以在等响曲线上方便地测查出其对应的响度级。

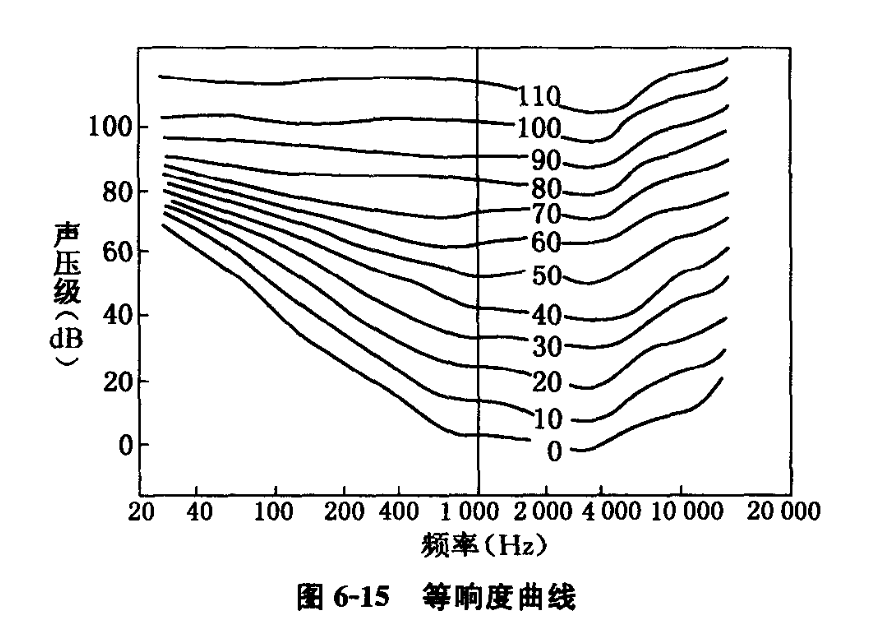


图 6-15 等响度曲线

图中最下面的一条曲线就是作为听阈的最小可听声强曲线,它是等响曲线的特例,它反映的是各种频率声音刺激的阈限声压级,故被称为零响度曲线。但是,必须注意的是,零响度曲线的响度级并不是 0 phone,而是 4.2 phone①,这是因为 1 000 Hz 的声音的听觉绝对感觉阈限是 4.2 dB。

三、听觉掩蔽的实验分析

听觉掩蔽(auditory masking)就是指由于某一声音刺激的存在而使另一声音刺激的强度阈限提高的现象,这是一种常见的听觉现象。比如,在安静环境里测试某被试对声音 A 的听觉绝对阈限值为 12 dB,同时在另一次实验中,由于被试在接受 A 刺激的同时,还接受到声音 B 的刺激,这时被试对声音 A 的听觉阈限发生了变化,声音 A 的声压级提高到 25 dB 才刚刚被听到,比原来的阈限值提高了 13 dB。这里的声音 B 起到了明显的干扰作用,故称为掩蔽音,而声音 A 就是被掩蔽音;25 dB 是有掩蔽音存在时对声音 A 的阈限值,称为掩蔽阈限;13 dB 是由于掩蔽音的存在,而使被掩蔽音听觉阈限提高的量,称为掩蔽量或掩蔽效应。听觉掩蔽反映了声音刺激识别之间的相互干扰,同时也可用于反映人耳对不同频率声音刺激的分辨能力,所以在实验心理学和实际应用中都获得了系统研究。

① 朱滢:《实验心理学》,北京大学出版社 2000 年版,第 206—207 页。

(一) 纯音掩蔽

听觉掩蔽的实验研究是从纯音掩蔽开始的,实验中以一纯音作为掩蔽音来掩蔽其他不同频率的纯音,并观察被掩蔽音感觉阈限的变化。

弗莱彻(Fletcher, 1953)的实验研究结果如图 6-16 所示。[①]图 6-16 中,A 图中的掩蔽音是 400 Hz,其强度包括 20、40、60、80 和 100 dB; B 图中的掩蔽音是 3 500 Hz,属于中高频,其强度也包括 20、40、60、80 和 100 dB 几种情况。在这个坐标系中,横坐标是被掩蔽音的频率,纵坐标是在掩蔽音影响下的掩蔽阈限。从图中显示的结果可以看出纯音掩蔽的以下主要规律。[②]

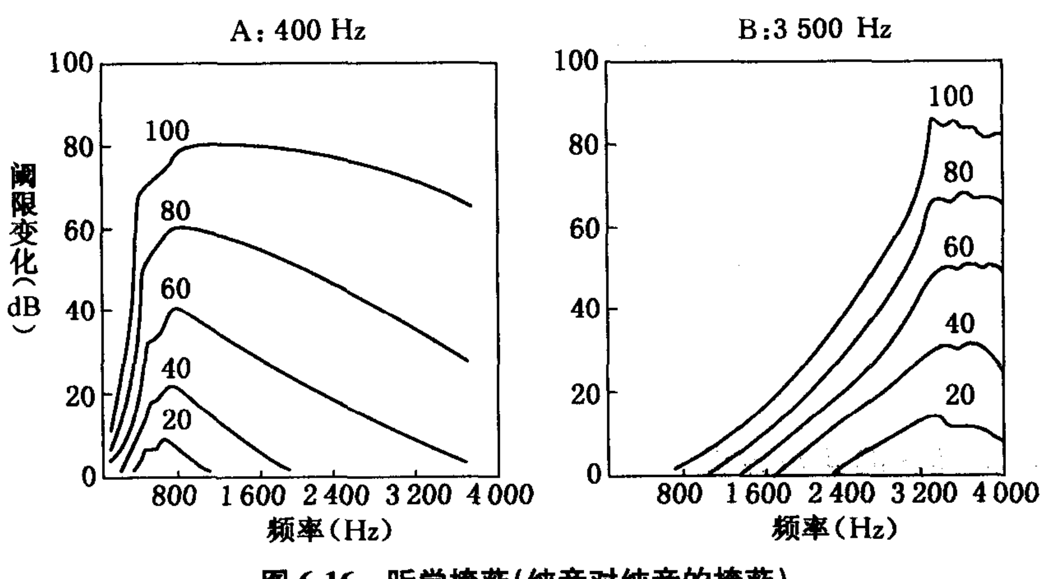


图 6-16 听觉掩蔽(纯音对纯音的掩蔽)

(1) 低频音的掩蔽效果更明显、掩蔽范围更大。低频音(400 Hz)不仅对低频音的掩蔽效果明显,造成其感觉阈限大幅提高,而且对中高频音造成了明显影响,使其感觉阈限大幅提高。但是,当以中高频音作为掩蔽音时,它对中高频音的影响很大,使其感觉阈限提高明显,对低频音的影响却很小。简单地说,就是一个声音引起的掩蔽主要决定于它的强度和频率,低频声能有效地掩蔽高频声,但高频声对低频声的掩蔽效果不大,因此,在应用听觉掩蔽时,不仅要关注掩蔽音的强度,还应关注掩蔽音和被掩蔽音的频率。

(2) 掩蔽音对于与其频率相近的声音的影响最大。如图 6-16 中的 400 Hz 掩蔽音对 200~1 000 Hz 纯音的影响效果明显大于对 1 000 Hz 以上纯音的影

---

① Fletcher, H. (1953). *Speech and hearing in communication*. New York: Van Nostrand.

② 杨治良:《基础实验心理学》,甘肃人民出版社 1988 年版,第 186—187 页。

响;3 500 Hz 掩蔽音对 3 000～4 000 Hz 纯音的影响效果也明显大于对 3 000 Hz 以下纯音的影响。

(3) 随着掩蔽音强度的提高,其掩蔽效果也迅速增加。如 400 Hz 的掩蔽音强度是 60 dB 时,800 Hz 的纯音强度要接近 40 dB 时才能听到;当该掩蔽音强度提高到 100 dB 时,800 Hz 的纯音强度也需要提高到约 80 dB 时才能听到。此外,掩蔽音强度越大,它影响的范围也越大,如 400 Hz 掩蔽音的强度为 20 dB 时,影响范围是 200～1 000 Hz,而当该掩蔽音的强度达到 80 dB 以上时,其影响的范围包括 0～4 000 Hz 及其以上的声音。

(二) 白噪声掩蔽

纯音对纯音的掩蔽实验中有一个不足之处,即拍频的影响。当掩蔽音和被掩蔽音的频率接近时,会产生一个拍声,这会降低掩蔽阈值。为了克服这一缺点,韦格尔和莱恩(Wegel & Lane, 1924)等进一步使用白噪声作为掩蔽音的实验研究,以考察其对纯音的掩蔽作用。实验结果显示,纯音产生在白噪声背景上时,会被白噪声掩蔽音所掩盖。随着噪声强度的增加,其掩盖的效果也明显增加,而且掩盖效果的增加与被掩蔽音的频率关系不大,主要由掩蔽噪声的强度决定,这一点与纯音掩蔽的情况不一样。

四、听觉告警的应用

听觉告警是人类社会生活中普遍存在的现象,它是利用相关设备通过声音向人们传送告警信息,以保障安全和防止事故。这些传送告警信息的装置叫做听觉告警显示器,①它的设计一直受到人类工效学研究人员和工程心理学家的特别关注。

(一) 听觉告警的优越性

听觉本身的特点和规律决定了听觉告警是最为优越的告警方式。与视觉或其他告警方式相比,听觉告警至少具有以下三个方面的优势。

1. *声音信号的全方位传送和接收*

告警信息可以通过多种感觉通道传递,包括视觉、听觉、嗅觉、味觉等方面,最常用的是视觉和听觉。视觉信息必须来自前方,才能被眼睛直接接收。左右

---

① 朱祖祥、葛列众、张智君:《工程心理学》,人民教育出版社 2000 年版,第 151—155 页。

两侧或背后的光线刺激不能被直接接收,只有当观察者转动头部时才能接收到。与视觉相比,听觉信息的传递和接收都可以是全方位的,一方面它在某一点发出,可以向各个方向发射,将告警信息传送到各个方位的人群;另一方面,人耳可以接收到来自各个方位的听觉信息,它不必要求告警显示器安装在正对着耳朵的方向,告警信息的传送和接收都是高效的。

2. *声音信号接收的被动性*

告警信息要求人们能快速接收并及时作出反应,但由于告警信息具有很大的不确定性,人们无法预知其发送的时间,所以不可能在告警信息即将发生时处在警觉状态,这就使得很多视觉告警信息不能很好地传达到相关人的眼睛中去。即使有时视觉告警信息能够进入眼睛,但也有出现"视而不见"的可能。但对于听觉信号来说,人们的接收具有被动性,一个人不管处在什么位置、身体保持什么姿势,声音只要能够以足够的强度或特殊的音调到达耳朵,就会引起人们的不随意注意,所以听觉告警信息一般都能被立即发觉,及时响应。

3. *声音信号可以穿透烟雾和绕过障碍物*

声音信号可以通过空气传播,能穿透烟雾、雨水、黑夜,能绕过围墙、隔板等障碍物,所以在许多不能使用视觉信号告警的情况下都能够使用听觉告警。在传输距离上,只要声音的强度足够大,也就可以将告警信息传送到相当远的距离,如城市中的防空警报。

(二) 听觉告警信号设计的工效学要求

1. *强度要求*

声音告警信号的强度,要视环境噪声水平和传送距离而定。告警信号的强度"应该比传送非告警信息的声音响一些。最好能保证信号比噪声强度超出8～15 dB的水平,关键是其响度水平要使相应范围内的人群都能接收到告警信息并引起注意。

2. *频率要求*

声音传播效能与声音本身的频率及传播通路的环境特点关系密切,因此在告警信号设计中要根据传播环境和传播要求确定信号频率。在多数情况下,听觉告警信号在500～3 000 Hz范围内,因为人耳对这一频率范围内的声音比较敏感,感觉阈限较低。具体情境中,要视具体情况而定:(1)传送距离超过300米以上时,最好采用1 000 Hz以下的告警信号,因为1 000 Hz以上的信号在传送中容易被吸收、衰减较快;(2)在有院墙等障碍物的场合下,最好使用500 Hz以

下的告警信号,因为低频音更容易绕过障碍物;(3)在噪声环境中,告警信号的频率应与噪声频率有较大差异,以降低噪声对告警信号的掩蔽作用。

3. 音色要求

告警信号一般选用音色特异的声音,以便容易引起人们的注意。音色主要由声音的波形决定,纯音的波形比较简单,音色特异性小,复合音往往具有较特殊的音色,所以告警信号通常都选用复合音。比如,波音飞机坐舱内使用某些动物的叫声作为紧急告警信号,其音色特异性强。

4. 对告警信号刺激模式的调制

为了增强信号的告警效果,设计者常常对声音信号的呈现模式进行编制,如音调的高低变化、强度的强弱变化、时间的连续和断续变化等。调制后的告警信号往往更能引起人们的注意和警觉。

根据声音传递的特点和规律,人们可以在特定环境中设计出适宜的听觉告警显示器,以起到应有的警示和预防保护作用。常见的听觉告警显示器有蜂鸣器、铃、哨子、钟、锣、号角、汽笛等,这些显示器发出的声音信号在强度、频率和音色等方面都有自己的特点,被应用于不同情境中的告警系统。

阅读材料 6-1

发现马赫带效应

如图 6-17 所示,在明度不等的两个面的相交处,出现了亮度对比,使双方各向着自己亮度的一方加强,即亮的一边更亮,暗的一边更暗。一百年前,奥地利

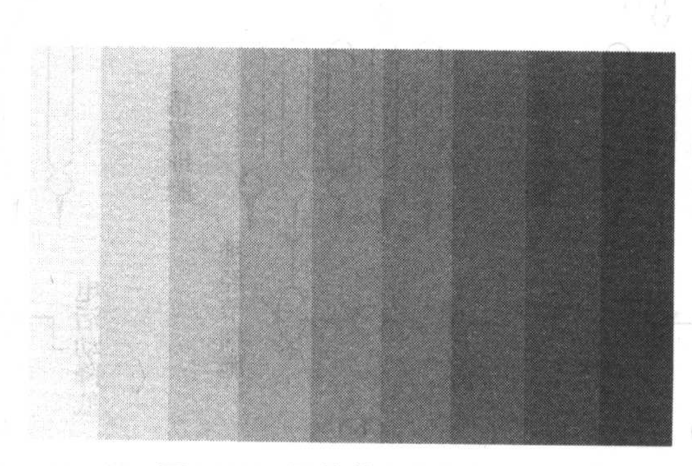


图 6-17 马赫带示意图

物理学家恩斯特·马赫(Ernst Mach)最早发现这种现象,于是就以他的名字命名,将这种现象叫做"马赫带现象"。

马赫最初以为这是一种纯粹的物理现象,然而他在进行大量的测量后发现,在同一条带子视觉上显示出的明暗两端并不存在物理刺激上的差异,而是均匀不变的。那么为什么人们看起来却总感觉不均匀而有明有暗呢?他认为这是一种心理现象。当他只让一条灰带显露,其余部分用纸遮盖起来,马赫带现象就不存在了。因此,他断定这是由于空间亮度对比造成的视错觉现象。马赫虽然天才地发现了这一现象,但他没有揭开这一现象之谜。

直到20世纪60年代,随着电生理学的发展,微电极技术的使用,在感觉系统神经元之间侧抑制的发现,才对马赫带现象作出了科学解释。1965年,美国学者瑞特利福(Ratliff)搜集并综合了大量有关的心理物理学和感觉生理学的研究成果,写成《马赫带》一书,生动而深刻地阐明了马赫带的实质。对这些成果加以概述,可进一步帮助我们理解马赫带效应。

1959年,美国的电生理学家哈特莱(Hartline)和康斯威特(Cornsweet)在号称活化石的鲎(hòu)的复眼中,用比头发丝还要细的微电极插入小眼的神经节细胞上进行探测,成功地发现了"侧抑制作用"。所谓侧抑制现象就是在相邻的神经元之间彼此抑制的现象。他们实验的简单模式是这样的:剥露鲎的复眼(由一千个小眼镶嵌组成),以光作为刺激,先用光单独照射某个小眼如A,可记录A小眼的神经节细胞的脉冲发放频率,如图6-18中a所示。而当光单独照射与A并排的另一个小眼B时,在A上没有记录到任何电反应,如图6-18中b所示。如果同时用光照射A、B小眼,就出现了侧抑制作用,无论记录其中哪一个小眼的电反应,

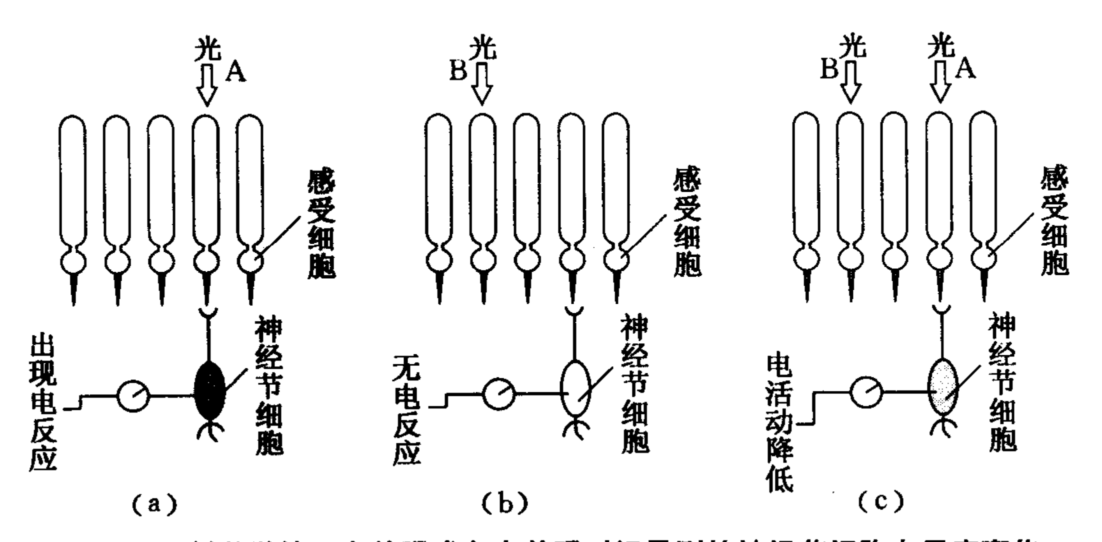


图6-18 刺激鲎的一个单眼或多个单眼时记录到的神经节细胞电反应变化

都可以看到脉冲发放频率降低的现象，如图6-18中c所示，这就是侧抑制作用的结果。侧抑制作用在人和动物的感觉系统内普遍存在，它的实现与神经元之间存在横向纤维网络有关，若切断了横向纤维，侧抑制作用也就不存在了。

侧抑制作用的大小与刺激强度的大小有关，又与神经元之间的距离有关。一般来说，刺激强度越弱，神经元之间距离越大，侧抑制作用就越弱。而且，侧抑制的作用是相互的。

由于神经元之间的侧抑制作用，使得视觉刺激强度出现阶梯的地方导致不同神经单元的反应出现了较大的反差。在人的视觉系统中，神经节细胞的大起大落，恰为大脑提供了夸张的信息。这就是马赫带现象中我们看到的两带交界处出现最亮与最暗感觉的生理机制。侧抑制作用的发现就在细胞水平上对马赫带现象作出了科学的说明。

人们认识了马赫带现象，可以自觉地加以利用，在生产、生活中发挥它的独特作用。如计算机的图形识别就利用了这一原理，用电子元件模拟人的神经细胞的功能，对主要信息进行加工。电视机中装有勾边电路，也利用了这一原理，使图像的轮廓勾画得突出、清晰而逼真。在绘画和摄影艺术上也利用人眼的这一功能，为突出绘画的艺术效果，在对象和背景之间涂色时采用烘云托月的手法，形成显著的差别，达到艺术效果。逆光摄影也利用马赫带效应的原理，使被摄对象的外缘披上耀眼的轮廓光，在深色的背景上烘托出主体，增强景物的空间深度感，有利于造型效果和艺术的表现力。马赫带效应的深入研究，将会在实践中发挥它越来越多的作用。

——资料来源：《心理学150问》(车文博，1987)①

阅读材料6-2

蝙蝠的回声定向

“方位定向(orientation)是指对物体的空间关系、位置和对机体自身所在空间位置的知觉。”动物和人都具有一定的方位定向能力和各自独特的方位定向方

① 车文博：《心理学150问》，辽宁人民出版社1987年版，第26—29页。

式。一般来说，人和动物的方位定向利用了多方面的感觉信息，对方位进行标记或判断。人主要利用了视觉和听觉两方面的知觉来把握方位，尤其以视觉为主。听觉方位判断则主要利用了两只耳朵接收的声音在音强、音色、相位等方面的差异。有趣的是，从某种角度来说，人未必是地球上方向定位能力最强的动物，除非他们使用了各种科学仪器。如信鸽能将信息传递到千里以外，然后准确无误地回到自己住处。

"蝙蝠的回声定向系统是自然界的最大奇迹之一。蝙蝠发出声音，声音冲击着环境中的物体，然后弹回给发声者，借助回音提供的信息，蝙蝠能避开飞行途中的障碍，并能定位、鉴别和捕获猎物。这是一种不可思议的成就。回声的振幅仅及发声振幅的千分之二，可是基于这种信息，蝙蝠就能判定在环境中存在的是猎物还是障碍。如果是猎物，连续的回声必须将关于猎物的方位、距离、速度及飞行路线的信息加以编码。对于蝙蝠定向性质的研究，至少可以追溯到18世纪意大利教士和科学家斯帕兰扎里(Lazara Spallanzani)。他发现实验的盲蝙蝠能绕着房间飞行而没有任何问题。朱赖利(Jurine，约1795)重复了斯帕兰扎里的观察，但得到了另外的惊人发现，如果用蜡封住这些盲蝙蝠的双耳，它们便失去了飞行能力。这些发现暗示听觉在蝙蝠的飞行能力中起着一定的作用。但是，直到1920年，哈特里格(Hartridge)才提出蝙蝠是利用它自己发出的超声波①的回声来指导飞行的。"

哈特里格的假设以后得到证实。后来科学家的进一步研究发现：不塞住嘴和耳的小蝙蝠能通过用细铁丝做的迷津；封住它的嘴和耳朵，它就没有这种能力了；如果只盖住它的双眼，它还具有这种能力。

现在初步认定，蝙蝠发出超声波和接收其回音是它的方位定向基础。令人感到不可思议的是，蝙蝠发出的声音是人们几乎无法听到的一种非常微弱的"咔嗒声"，这种声音怎能有足够的能量冲击障碍物并通过回音提供精细的信息呢？这些问题是现代仿生学正在探索的问题。

"蝙蝠的回声定向系统提供了高度复杂的、适应性的知觉系统。回声定向系统使这类动物对猎物的距离和方向能作精密的判断。此外，它还提供了洞穴中飞行时精细调整所需要的信息，以免碰撞墙壁、岩石或其他物体。"

——资料来源：《普通心理学》(彭聃龄，2001)②

① 超声波就是高频率声波，对人来说，就是超出可听声波范围的20 000 Hz频率以上的声波。

② 转引自：彭聃龄：《普通心理学》，北京师范大学出版社2001年(第二版)，第152—153页。

建议阅读文献

1. 杨治良:《实验心理学》,浙江教育出版社 1998 年版,第 273—394 页。
2. 北京师范大学等四校联合编写:《人体解剖生理学》,高等教育出版社 1982 年版,第 111—132 页。
3. 李新旺:《生理心理学导论》,河南大学出版社 1992 年版,第 108—163 页。

复习思考题

1. 如何理解可见光、光通量、光照度、光亮度、光锥、CIE 色度分析系统、视敏度、时间视敏度、空间视敏度、闪光融合临界频率、暗适应曲线、三色说、色盲、色弱、噪声、白噪声、等响度曲线、零响度曲线、分贝、sone、phone、听觉掩蔽、语图仪?
2. 简述眼睛的结构。
3. 简述视网膜的结构及其对信息的接收、加工与传递过程。
4. 简述颜色视觉形成的机制。
5. 如何测量视敏度?
6. 如何测定和制作暗适应曲线?
7. 暗适应的影响因素有哪些? 其在实际中的应用有哪些?
8. 简述耳的结构及其接收、传递信息的机制。
9. 听觉掩蔽有什么特点和规律?
10. 音高感是由什么因素决定的?
11. 响度感是由什么因素决定的?
12. 如何测定和制作等响度曲线?
13. 听觉告警的优点有哪些? 听觉告警显示器的设计有什么要求?

第七章
空间知觉的实验研究

本章内容提要

知觉是人类最基本的认知过程，它是对感觉信息的组织与解释。关于知觉的形成机制，不同心理学家有完全不同的看法，目前存在三种典型理论：直通知觉理论、建构知觉理论和计算知觉理论。在具体的知觉过程中，又存在各种不同的信息加工方式，主要包括：自下而上的加工或数据驱动的加工、自上而下的加工或概念驱动的加工。在知觉信息加工过程中，存在较为普遍的整体优势效应，如字词优势效应、客体优势效应等。从内容来看，知觉包括物体知觉和社会知觉两大类，社会知觉就是关于人的知觉，它主要是社会心理学的研究内容；物体知觉则主要包括空间知觉、时间知觉和运动知觉三方面。由于迄今关于时间和运动知觉的研究积累相对较少，本章主要讨论有关空间知觉的一些基本研究和相关的经典实验。空间知觉主要是关于物体的大小、形状和距离的知觉，这里讨论形状知觉、深度知觉和大小知觉恒常性等问题。

知觉能力是人的基本认知能力和生存能力，因此知觉(perception)一直是心理学基础研究的重要领域。这一领域首先要阐明的是知觉的本质和知觉发生的机制，然后将研究扩展到某一具体分支，探讨知觉活动的特定方面。基于知觉研究的发展历程与当前趋向，我们首先介绍当代认知心理学中有代表性的知觉理论，然后对空间知觉(spatial perception)的发生进行实验分析。

第一节　知觉理论及其实验验证

"知觉是按一定方式来整合个别的感觉信息，形成一定的结构，并根据个体

的经验来解释由感觉提供的信息。”①它作为一种基本的心理活动，包含了互相联系的几种作用或过程：觉察（detection）、分辨（discrimination）和确认（identification）（Moates, 1980）。觉察是指发现事物的存在，但尚不清楚它是什么；分辨是把一个事物或属性与其他事物或属性区分开来，也就是知觉对象与背景相分离的过程；确认是指人们利用已有知识经验和当前获得的信息，确定知觉的对象是什么，给它命名，并把它纳入一定的范畴。在这一知觉过程中，还有假设检验的成分，而且知识经验在其中起到不可缺少的作用。

通常，知觉是快速短暂的，我们甚至意识不到它的细节，而只是意识到它的结果。但是当感觉信息不充分、知觉条件不佳、知觉过程延续时间较长时，就可以意识到知觉过程，有时这种意识相当清晰，以至于可以用言语描述。比如，图7-1所示的刺激物是什么呢？人们会在知觉过程中分析它的特征、作出假设、赋予它一定的意义，但随即又被自己否认。

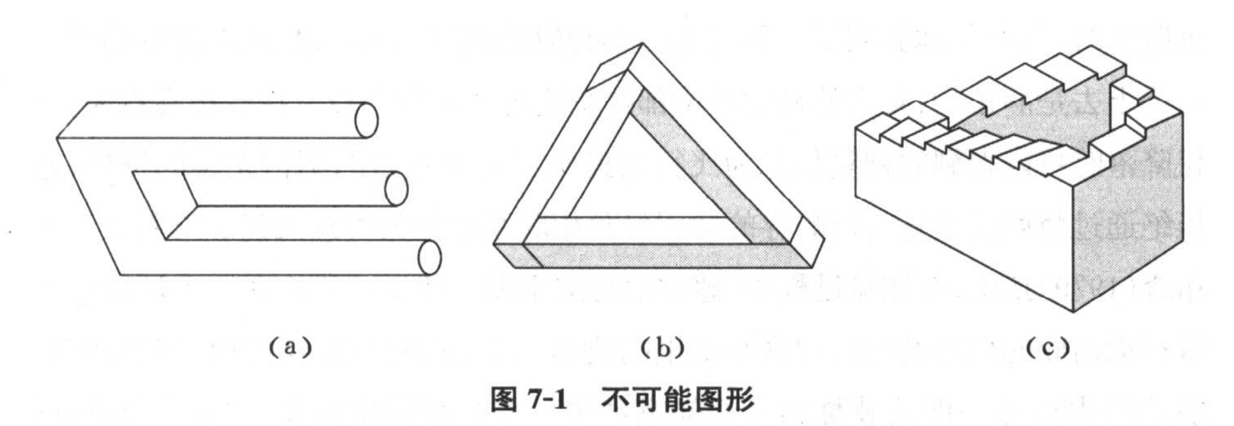


图7-1　不可能图形

现代实验心理学对视知觉进行了大量研究，虽然不同学者对知觉的精细结构和过程有着非常不同的看法，但有一点是非常确定的，那就是知觉过程具有明显的组织性，它能将分离的、局部的结构整合在一起获得一个良好的解释，其中表现出对刺激材料的组织、假设、补充。这一点从图7-2所示的“主观轮廓”知觉现象能获得清楚的说明。在观察这一图形时，视野中的明度分布能构成的最直接视觉印象应

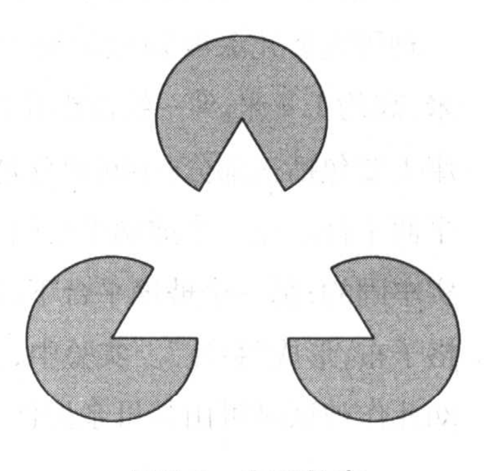


图7-2　主观轮廓

① 彭聃龄：《普通心理学》，北京师范大学出版社2001年版，第125页。

该是三个残缺的黑色的圆盘,或者说是三个扇形,除此之外再无别的。我们却分明看到了一个白色不透明的三角形盖在了三个黑色的圆盘上,三个黑色圆盘也不被认为是残缺的,而是被白色三角形部分地盖住了。

知觉印象是如何形成的呢? 这是知觉研究领域一个带有根本性的问题,也是该领域争论的焦点,对此有不同的理论加以解释。我们这里介绍其中具有代表性的三个理论:直通知觉理论、建构知觉理论和计算知觉理论。

一、直通知觉理论

直通知觉理论也叫直接知觉理论(direct perception theory),它主要以吉布森(James J. Gibson)的研究为基础。第二次世界大战期间,吉布森服务于美国某空军基地,他的工作与飞机降落的问题有关。在研究中,他发现当飞机降落而接近跑道时,飞行员体验到了一种与光流相类似的视觉幻象:飞机正在接近的那一点看上去是静止的,其余的视觉环境都在快速地离开那一点。在吉布森看来,飞机降落时可以得到的视觉信息为飞行员提供了关于跑道的明白无误的资料,他拒绝通过将刺激信息与已贮存的知识经验相联系而把某种意义强加给刺激。吉布森(1979)指出,在知觉过程中,感觉信息更重要。他说:"我们在一个环境中采取行动需要的信息统统都包含在那里的物体当中,我们不需要任何别的知觉机能。"①因此,这一理论被称为直通知觉理论。吉布森还使用关于"视崖"(visual cliff)的实验研究来支持这一理论。视崖是吉布森等人在20世纪60年代设计的一种研究婴儿深度知觉的装置,如图7-3所示。该装置长约2.6米,宽约1.3米,高约1.4米。②一般由钢化玻璃作为支撑物,婴儿在上面爬行是安全的。视崖主要包括三部分:中间部分是一木质平台、两头使用钢化玻璃搭建成两个玻璃平面平台。在一个玻璃平台的下面紧挨着玻璃贴上格子布,看上去其下面是充实牢固的;另一个玻璃平台下面则做成空箱,并且在空箱的底部及一个侧壁贴上格子布,形成"悬崖"。实验中,将幼儿或小动物放在中间实体平台上,主试(使用幼儿作被试时可由其母亲担任)站在仪器一端,呼唤被试或通过食物诱导被试的

① Gavin, H. (Ed.)(1998). *The essence of cognitive psychology*. London: Prentice Hall Europe, 26.

② 林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社2003年版,第1152—1153页。

行为，同时记录被试爬向主试的次数和每次“犹豫”的时间。通过视崖实验，吉布森等人发现六个月的儿童就不会或很少爬上有“悬崖”或看上去容易“跌落”(drop)的那一边去，即使这些孩子没有过掉下悬崖的任何经历。吉布森解释说，接受深度暗示和躲避跌落是人的先天智慧。

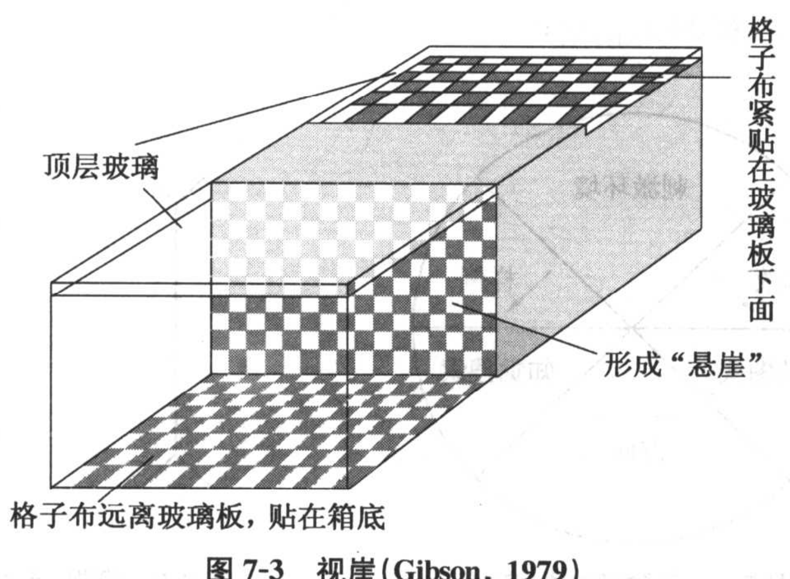


图 7-3　视崖(Gibson，1979)

在正确知觉的情况下，吉布森的理论可以给予知觉现象一完整的解释；但在发生错觉的情况下，吉布森的理论就无法给出合理解释。于是其他心理学家提出了与之完全不同的观点。

二、建构知觉理论

格雷戈里(Gregory, 1972)认为知觉是对感觉碎片和来自记忆的信息碎片进行整合的过程，这一观点被称为建构知觉理论(constructive perception theory)。依此理论，错误的知觉和不完整的知觉都容易得到解释。在有些情况下，刺激信息本身是不充分的，容易导致错误的知觉建构。

奈瑟(Neisser, 1967)在进一步研究的基础上提出了知觉中包括图式(schema)、搜索(exploration)和刺激(stimuli)的观点。所谓知觉图式(perception schema)，就是将过去感知过的一些刺激信息有机结合在一起构成的一个结构稳定的内部表征。知觉图式可在刺激作用下被激活，然后认知主体依靠它就可以对不充分的信息进行预期，并指导对环境刺激信息的进一步搜索，如图 7-4 所示。搜索意味着多方向的搜索运动，并接受可利用刺激构成的样例(sample)，当样例

与内部图式不匹配时，图式就有可能被修改。如果知觉的刺激质量很差，那么图式就要起到更大作用。建构知觉理论在解释视错觉方面优于直通知觉理论。如果已贮存的信息能够被用于解释当前刺激的意义，那么错觉可能就是错误地应用这些知识的结果。那么，利用建构知觉理论能否解释缪勒-莱尔错觉（Müller-Lyer illusion，简称 M-L 错觉）？

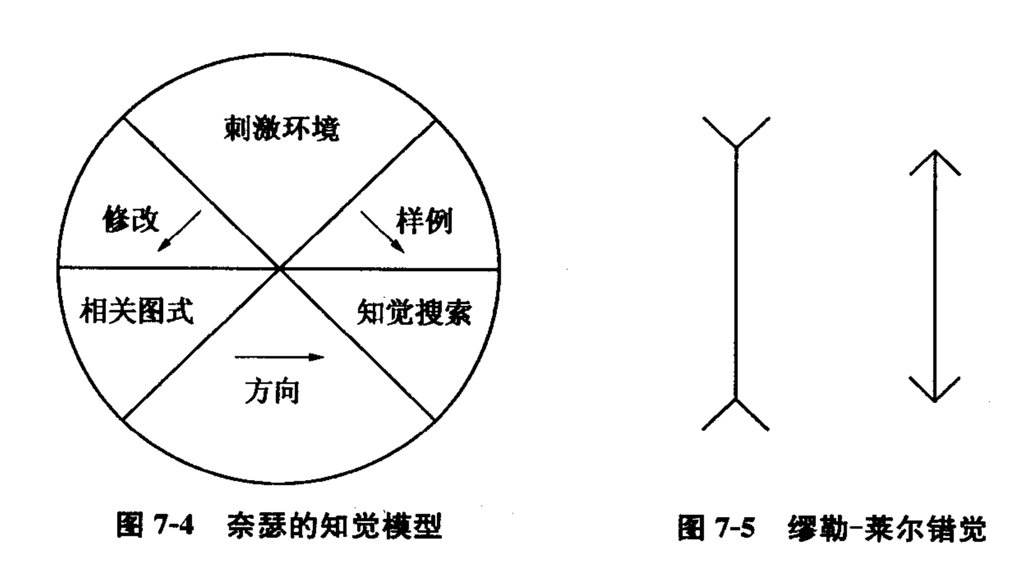


图 7-4　奈瑟的知觉模型　　**图 7-5　缪勒-莱尔错觉**

格雷戈里（Gregory，1972）的解释是：我们会在接受这样的刺激时激活贮存的图式，进而把图 7-5 中左边的线条看成是一个房间的内角、右边的线条看成是一个建筑物的外角。换句话说，图中左边的线条好像正离我们远去，因此它被觉察为离我们更远些。根据大小恒常性规律，我们会把那些能形成相同网膜像的不同物体中离我们越远的知觉为越大。图 7-5 中两条线段的实际长度一样，所以在网膜上形成大小相等的像，当把左边线段知觉为较远时，它也就会被知觉为越长了。这就是知觉的恒常性误用，也是稳定的心理图式对知觉过程发生影响的一个例证。

三、计算知觉理论

计算知觉理论（computational perception theory）是马尔（Marr，1982）在对人的知觉过程进行计算机模拟后提出来的。马尔根据信息加工观点，把视知觉看作是一系列复杂、连续的加工阶段或步骤，在此过程中的每一阶段都要从外部世界抽取信息并将其加载到此前建立起来的视觉模式上去，知觉过程中的每一阶段都包含大量的计算细节。通常，人们知觉到的是计算的结果，而不是计算的

过程。可以设想，如果我们在知觉一个复杂图形时需要许多步骤的计算过程，在不同计算阶段得到的计算结果也应该是不一样的，所以在不同时刻将一个较复杂的知觉过程中止，就可以在不同时刻获得不同的知觉结果。比如，给被试呈现一个 20 ms 的瞬时红光刺激，紧随其后在同一位置上呈现一个 20 ms 的绿光刺激。被试报告看到了“黄色”的闪光刺激，就如同红光和绿光同时闪烁时看到的情形一样。不过，如果绿色闪光不是紧随红光之后而是间隔了一定的时间，被试就会报告先是看到红色闪光，然后看到的是绿光刺激。这说明，直到来自绿光的信息被加载完成之前，被试不可能意识到黄光的存在。在整个知觉过程中，被试是先接受红光刺激，紧接着又接受绿光刺激，并且把绿光刺激的能量加载到了前 20 ms 的知觉结果上去。如果绿光延迟呈现，红光刺激就可以作为一个独立的刺激形成完整的知觉印象。看来，知觉有一个时间阈限值，在这个阈限内，加载的新刺激都会被计算进去参与生成一个完整的知觉印象。美国心理学家罗伯特·雷诺兹(Robert Reynods)对此进行过更详细的研究。他希望能够说明，知觉的不同方面可以在不同时刻被看到，换句话说，他试图探明从刺激呈现到形成相对稳定的知觉的时间历程。他的研究证明了知觉计算需要一个较长的时间周期。比如，我们如果还利用上面的主观轮廓图形做实验的话，就可以这样来控制实验过程：当刺激呈现 50 ms 时立即呈现一延迟刺激(掩蔽刺激，masking stimulus)，该刺激的模式更大更亮，被试就能清晰地看到三个有缺口的圆盘，但大部分被试都看不到幻觉三角形，少数被试报告说看到了幻觉三角形但不是很清晰、较难确定；如果掩蔽刺激延迟到 75 ms 呈现，则所有的观察者都报告说看到了三角形。这一实验结果可以解释为，知觉的起始阶段是对刺激能量的直接接收和计算，而在计算基础上的推断则是知觉的后续阶段。知觉计算过程持续的时间大致在 60～70 ms(Robert Efron, 1967)，在此周期内连续出现的刺激都会进入一个计算过程参与生成一个完整的知觉结果并被意识到，因此 20 ms 的红光刺激与紧随其后的 20 ms 的绿光刺激构成一个计算结果，即生成黄光知觉印象，而不是先生成红光知觉结果、再生成绿光知觉结果。1887 年，法国科学家查蓬特尔(Charpentier, A.)发现，长达 66 ms 的闪光刺激，看起来并不比 7 ms 的闪光刺激持续更长的时间。不过，这个数字是对较容易观察的突出刺激而言的。对于那些不清楚或较为复杂的刺激来说，其知觉计算的时间会更长。

上述三种理论各有其特点和优缺点。直通知觉理论反对知觉过程的分析性解释，强调环境刺激的作用和知觉的现实真实性，能够解释我们可以观察得到的

一些知觉现象,但是它对知觉的解释浮于现象层面,不能深入,因而不能为模式识别和人工智能开发提供多少有价值的资料。建构知觉理论则强调了环境刺激信息与已有知识经验的相互作用,充分体现了人的主观能动性,而且在解释知觉现象方面更具灵活性,但这一理论对知觉的细微过程描述不清,包含更多假设的成分。计算知觉理论,则是建立在毫秒单位时间维度分析基础上的精细理论,既不否认知觉对于环境刺激的依赖,也不否认知识图式在知觉中的作用,包含假设和推断的内容,因此是新近出现的很有研究价值的知觉理论。可以认为,计算知觉理论的提出将会推动知觉研究的不断深入和精细化,不仅能够为我们理解知觉现象提供帮助,也将为模式识别和人工智能开发提供新思路。而且,借助计算机技术,一般的认知实验室都可开展此方面的实验研究。简言之,计算知觉理论或许可以为知觉的精细研究开辟一条新道路,而且这条道路可以充分利用现代计算机技术以及精巧的实验技术。

第二节　知觉加工方式的实验研究

知觉加工的方式是关于知觉过程的概括性表述,它讨论知觉加工的主要特点和一般规律。

一、知觉加工的方式

为了说明知觉加工的方式,首先分析“瞎子摸象”的故事:摸到大象腿的瞎子说大象像一根柱子,摸到大象耳朵的瞎子说大象像一把大蒲扇,摸到大象尾巴的瞎子说大象像一根绳子……瞎子对大象的知觉取决于他触摸到的部分的特征,是刺激信息输入和组织的结果,这种加工就是由低层次的感觉信息上升到知觉组织的高级过程,所以被称为自下而上的加工(bottom-up processing)过程。因为这种知觉是刺激资料驱动的结果,所以也叫做资料驱动的加工或数据驱动的加工(data-driven processing)。

然而,我们也能很容易地举出例子来说明,知觉加工不总是按从搜索刺激材料到组合材料的顺序完成的。在许多情况下,人们可以依靠对背景信息的知觉而对刺激的整体先形成一个概念,然后在这个概念指导下对其中的部分进行知觉和预期,明显表现出已有知识经验或概念在知觉中的重要作用。认知心理学

教科书常常引用的一个研究是沃伦等人(Warren et al., 1970)进行的音素恢复实验。当给 20 名被试听下述句子时,句子中星形位置的音素被一个持续 120 ms 的纯音取代。在全部被试中,只有 1 人说他听到了该纯音,且不能指出该纯音的位置,而其他所有被试都没有发现 s 字母的缺失。

The state governors met with their respective legi * latures convening in capital city.

研究者进一步将此实验扩充,让不同的被试分别听一个不同的句子。所用的句子如下:

It was found that the * eel was on the axle.
It was found that the * eel was on the shoe.
It was found that the * eel was on the orange.
It was found that the * eel was on the table.

在每个句子中,星形仍然表示该位置上的字母和音节缺失。结果发现,听第一个句子的被试倾向于将缺失一个字母的词听成 wheel,听第二个句子的被试倾向于将其听成 heel,听第三个、第四个句子的被试分别倾向于将其听成 peel 和 meal。这一实验结果表明,被试在知觉某个刺激信息不完整的单词时受到了整个句子语义情境的影响,即被试先形成了对整个句子的知觉,形成了某种概念性理解,然后依此概念性理解指导了对个别单词的知觉,即使这个单词的刺激信息不完整。所有的被试并没有从这个单词本身的构成方面来知觉这个单词,因为那样的话,所有的被试就会得到相同的知觉,即"eel"音节。

这一实验显示的先形成概念性理解之后再在概念指引下的知觉方向,是从抽象性更高的层次指向抽象性较低层次的知觉加工,故叫做自上而下的加工(top-down processing),也叫做概念驱动加工(concept-driven processing)。在知觉过程中的概念性指导来自过去的知识经验,有知识经验参与的知觉过程好于无相关经验或相关经验较少的知觉过程,比如米勒和伊沙德(Miller & Isard, 1963)的实验。

米勒和伊沙德在噪声背景上,给被试听一些句子,要求被试将听到的句子报

告出来。这些句子分成三类:第一类为正常句子,既合乎语法,又有正常语义;第二类为异常句子,虽然合乎语法但语义不合理;第三类句子为非语法句,既不合乎语法也无合理语义。如:

第一类句子:

A witness signed the official document.

Sloppy fielding loses baseball games.

第二类句子:

A witness appraised the shocking company dragon.

Sloppy poetry leaves nuclear minutes.

第三类句子:

A legal glittering the exposed picnic knight.

Loses poetry spots total wasted.

这三类句子被正确知觉的百分数有所不同,如图 7-6 所示。

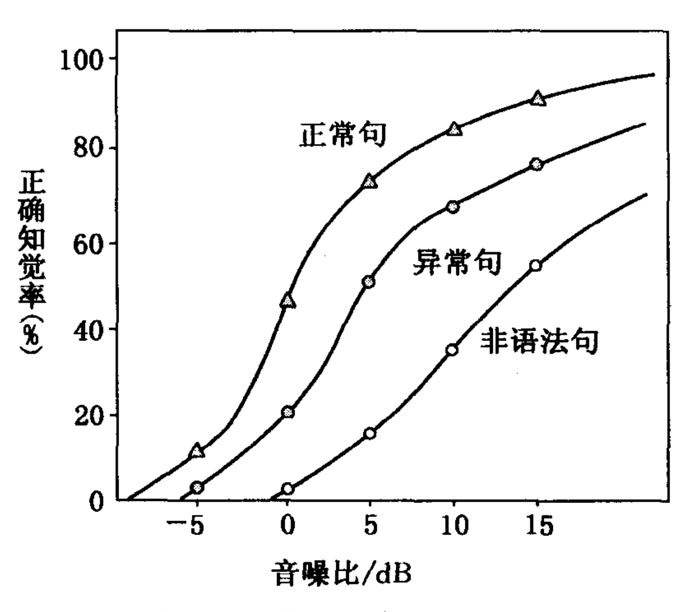


图 7-6 句子知觉是句子类型和音噪比的函数

概括上述实验,可以根据知觉方式是按照由具体散件的刺激到刺激整合的方向,还是按照从一般知识和概念到具体部件的方向分为两大类:数据驱动的加工或自下而上的加工、概念驱动的加工或自上而下的加工(Lindsay & Norman, 1977)。前者指从刺激作用开始的加工,一般先对较小的刺激单元进行分析,进而转入较大的刺激单元,通过一系列连续加工过程,最后完成对引起感觉的刺激信息的解释;后者往往是从人的期望或对于知觉对象的假设开始的,并对信息加

工的各个阶段和水平加以制约，其中包括调整感觉器官对刺激信息的选择和引导对特定细节的注意等活动，如阅读时对于上下文关系的理解、对于双关图形的理解。知觉中言语暗示的明显效应也是概念驱动加工的典型例证。

上述分析涉及的是知觉加工中的时间特性，那么知觉加工的空间特性如何？就是说当同时呈现的刺激单元在空间有一定的排列关系时，知觉加工是按照一定的空间顺序进行加工，还是同时对各个不同空间位置上的刺激同时进行加工呢？这就是系列加工(serial processing)和平行加工(parallel processing)的方式。所谓系列加工就是对不同信息单元逐项、逐步地加工，比如我们在精读教材时就是按照文字排列的先后顺序逐字逐句地读，并且在阅读每个字词时都是从接受其刺激、构造其字形、识别其字意、理解其词义和句意等，这就是系列加工。但有时，我们为了快速搜索某一特定的信息，可以“一目十行”，同时加工不同的信息单元，这就是平行加工。所谓平行加工，就是指多方面的刺激信息可以同时在不同的加工单元中被处理，这种加工方式可以提高信息处理的速度和效率。

二、知觉加工中的优势效应

知觉过程受多方面因素的影响。从前文的讨论可以看出，知识经验的影响是明显的。此外，格式塔心理学非常强调“完形”(Gestalt)的价值，这就是强调了知觉过程中的整体优势效应。对于这一点，当代认知心理学中有一系列相关研究，如结构优势效应、字词优势效应、客体优势效应等。

(一) 整体优势效应

在对客体进行知觉时，究竟是先知觉它的部分，后知觉其整体；还是先知觉其整体，后知觉其部分？研究表明，在一般的知觉活动中，既有整体加工(global processing)，也有局部加工(local processing)。整体加工就是知觉到刺激物的整体特征(global feature)的加工，局部加工就是知觉到刺激物的局部特征(local feature)的加工。通常情况下，整体加工优先于局部加工。例如在一项测试中，要求被试对一些由各种小字母组成的大字母予以辨认，并测定其反应时间。结果表明，整体特征的知觉所需要的时间少于局部特征的知觉时间(Navon, 1977)。但是，另有一些研究结果显示，当小字母的刺激达到 8 度视角时，其知觉时间短于大字母的知觉时间(Kinchia & Wolfe, 1979)。看来，在视知觉过程中优先加工的是那些有着最佳大小的刺激物。那么，整体知觉究竟是否具有优势效应？下面介绍一些具体研究。

（二）字词优势效应

字词优势效应(word-superiority effect)是由雷彻(Reicher)在1969年的一项实验中发现的。他准备了三类视觉刺激材料(卡片)：第一类是只有一个或两个字母；第二类是一个或两个由四个字母组成的单词；第三类是一个或两个由四个字母组成的无意义字母串，如图7-7所示。

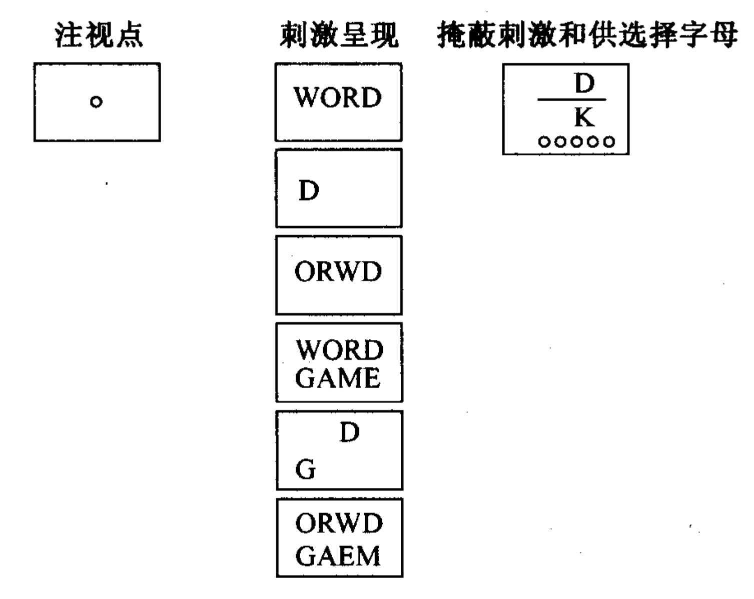


图7-7 雷彻的实验程序和刺激材料

上述所有刺激材料在实验中均使用速示器呈现给被试。在每一轮实验中，要求被试先注意速示器中出现的一个注视点，然后按事前拟定的程序在规定的短时间内呈现某一种刺激材料，随后立即换成掩蔽刺激和供被试选择的两个字母。这两个字母所在的位置恰好对应于刺激材料中需要测试的字母位置。每次实验后，要求被试选择最后出现的两个字母中哪个是前面刺激材料中在相同位置上出现过的，研究者根据被试选择的结果按对或错予以记录。当刺激材料为一个字母、一个单词或一个无意义的字母串时，提供选择的两个字母位于刺激材料中相对应字母位置的上方；而在使用其他刺激材料时，则两个供选择的字母中有一个在下方。刺激呈现的时间按照每个被试反应达到的正确率所需要的时间而定(以ms计)，分为短、长、中三种，其相应的时间分别为正确率达到60%所需时间、达到90%所需时间加上5 ms以及这两个时间的中点值。由此实验可见，每个被试在观察同样的刺激时所需要的时间是不同的。另外，实验的处理方式又分为两种：一种有先行信息，即在每次实验之前就把将要提供选择的两个字母告诉被试；还有一种则事先不告诉被试。研究结果如图7-8所示，不管有无先行

信息，对单词中的一个字母的识别成绩均优于对一个字母或无意义字母串的识别成绩。其识别的正确率大约相差 8%，并在统计学上达到显著性水平。

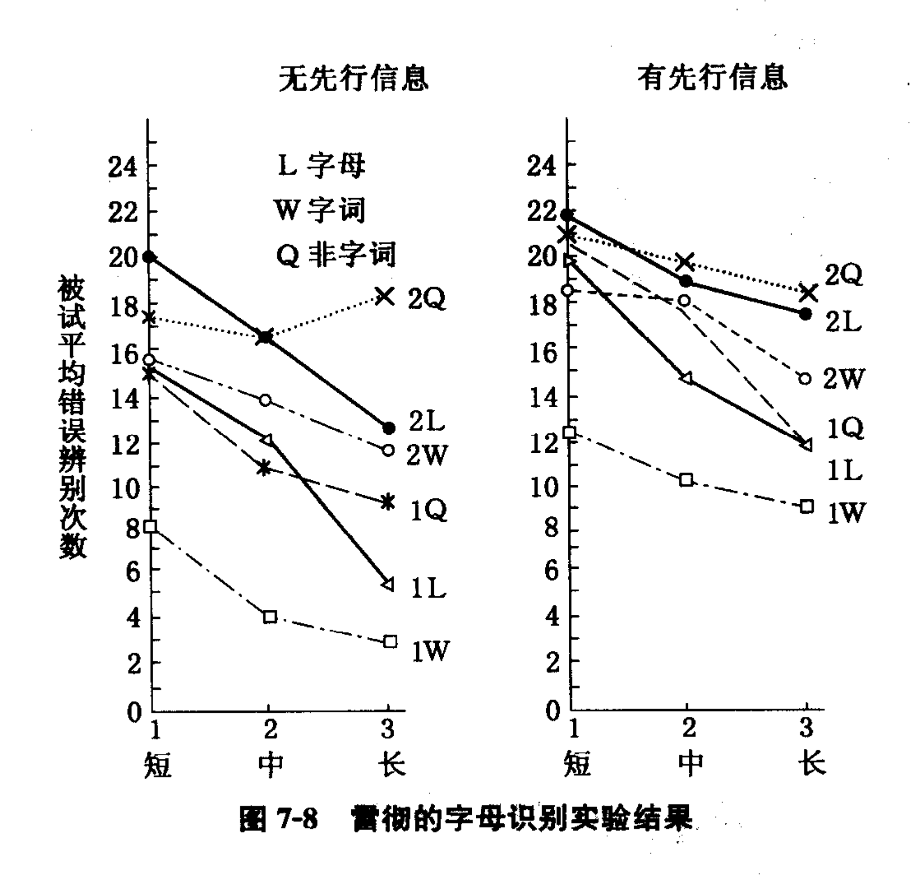


图 7-8 雷彻的字母识别实验结果

雷彻的研究引起心理学界的广泛注意以及许多研究者的兴趣，促进了这方面的一系列实验研究，并证实和肯定了字词优势效应。不仅如此，大量的研究还积累了丰富的资料，产生了各种有关字词优势效应的解释，其中有强调知识经验为基础的自上而下加工的推论说，也有以信息在内部加工而成的内码表征为主导的编码说，还有的理论则依据整体加工和局部加工的关系进行分析和解释。无论哪一种解释，都有一定的合理性，这说明人的知觉过程需经历多个不同的阶段和水平，只有把各种解释的合理成分综合起来，才可能对知觉中的复杂规律性有较全面的正确认识。

(三) 客体优势效应

字词优势效应研究的成果推动了心理学家们开展有关图形中线段识别的研究工作。早在 1974 年，韦斯坦和哈里斯(Weisstein, N. & Harris, C. S.)就进行了对包含在各种图形中的直线线段的识别研究。刺激材料是通过显示器呈现给被试的。在这些刺激材料的图形中，包含有要求被试识别的“靶线段”(target line)和与之有关联的关系图形(或称为上下文图形)，如图 7-9 所示。

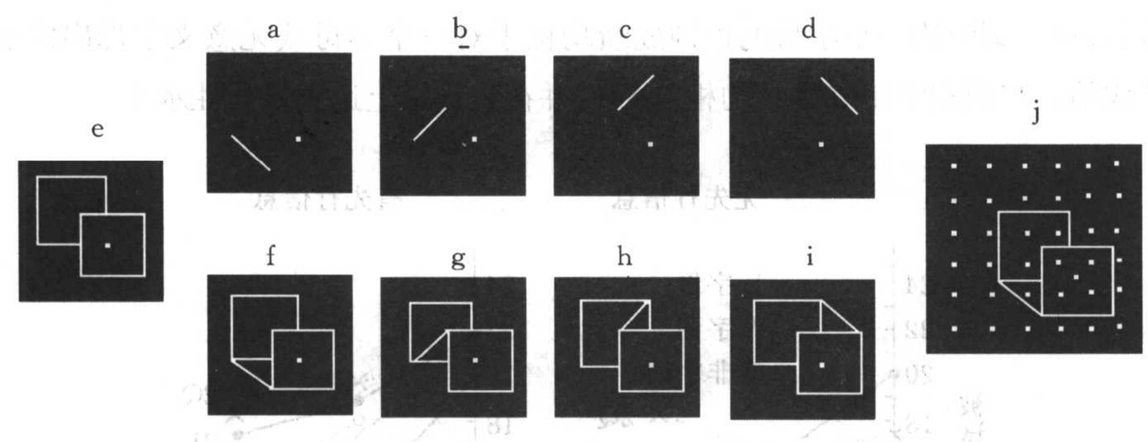


图 7-9　韦斯坦和哈里斯应用的刺激材料举例

在图 7-9 中各卡片上的白点是注视点。作为“靶线段”的刺激材料共有四种，它们除了在方向和位置上不同外，都具有相同的长度和宽度，图中所示的 e 为关系图形，f、g、h、i 为实验中呈现给被试的刺激图形的部分样例，j 是在刺激图形关闭后呈现的掩蔽图形。实验中实际使用的刺激图形，在三维结构的完整程度上分成六个等级，如图 7-10 所示。

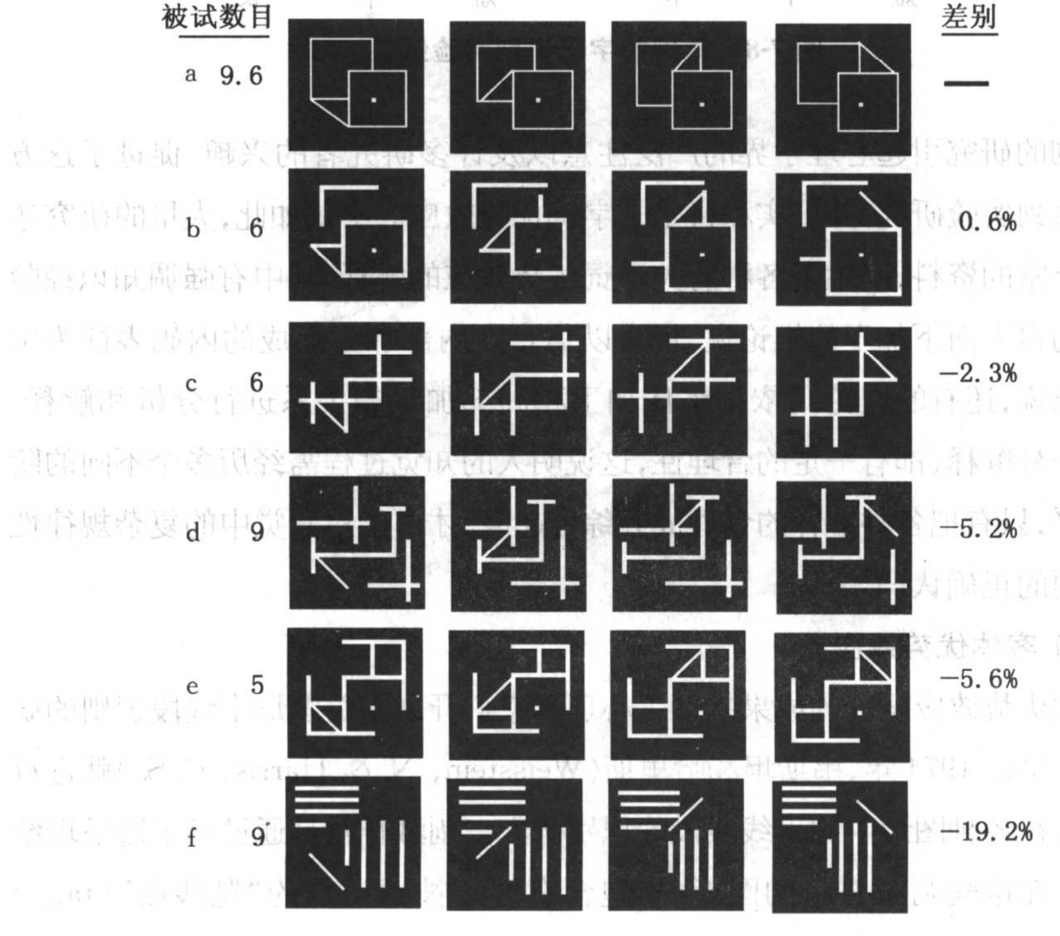


图 7-10　韦斯坦和哈里斯应用的各种上下文图形及其实验结果

以上所有图形在呈现时都有一固定注视点,要求被试把视线对准它。实验在暗室里进行,所有图形都在显示器上呈现。先让被试分别熟悉上述四种不在关系图形中并单独呈现的靶线段,然后由被试自行操作进入测试阶段。这时,显示器上先呈现一个注视点,被试在确认已经准备好的情况下,启动开关后立即呈现一个由靶线段和关系图形相结合的刺激图形(图 7-10 中的一个)。在刺激图形呈现后,随即换成一个呈现时间为 100 ms 的掩蔽图形(如图 7-9 中的 j)。这时,要求被试判断在刺激图形中出现的是四种靶线段中的哪一种,并按照规定在某个按键上予以反应。主试将判断的正误告知被试。如同上述字词优势效应的实验一样,刺激图形呈现的时间根据被试判断能达到的不同准确率(分别为 50%、60%和 70%)而定为三种,其跨度在 5～44 ms。实验程序按完全随机呈现设计。结果表明,位于一个三维结构完整和严谨的刺激图形内的靶线段识别率明显高于它们在那些结构性较差的刺激图形中的结果,这被称为客体优势效应(object-superiority effect)。

从某种意义上说,字词优势效应和客体优势效应都是整体优势效应的具体表现,整体结构的合理有序是其中的关键因素,它容易在知觉过程中激活有关的知识经验,快速对刺激模式予以解释和预期。在预期的引导下,可以形成对局部刺激或部件的定向,也容易对之识别,所以整体优势效应表现出明显的概念驱动性质。

第三节　形状知觉与图形后效实验

形状知觉(shape perception)是视知觉的最基本内容,也是个体获取知识经验(如观察物体的外形、阅读和看图等)最重要的途径之一。形状知觉研究首先要探明的是形状知觉形成的过程,包括它与轮廓的关系问题,因为轮廓为形状知觉提供了基本的信息载体,是形状知觉的条件。

一、形状与轮廓

(一) 轮廓

轮廓(outline)是物体的外形线。当视野中的明度或色调发生突变而将视野划分成不同的区域时,就有了轮廓。图形或形状是由一个可见的轮廓从视野中

分割出来的一块面积。①轮廓与形状相辅相成但又不同,视野中的两个部分被轮廓分开时,虽然共用一个轮廓,但仍具有不同的形状。

加里和霍奇海默(Galli & Hochheimer)在一项实验研究中,用一块黑色长方形纸板,沿着不规则的曲线切分为二,再将此二部分各自贴在白色纸板上,分别用速示器呈现给被试,使轮廓曲线处于被试的正前方的位置。每一次呈现后都要求被试用铅笔尽可能重复画出这个轮廓。因为,每一个轮廓会随同左边和右边的图形各呈现一次,但是被试却很难认出它们是同一个轮廓。这一情况还可以用图 7-11 所示的轮廓与图形的关系来说明。

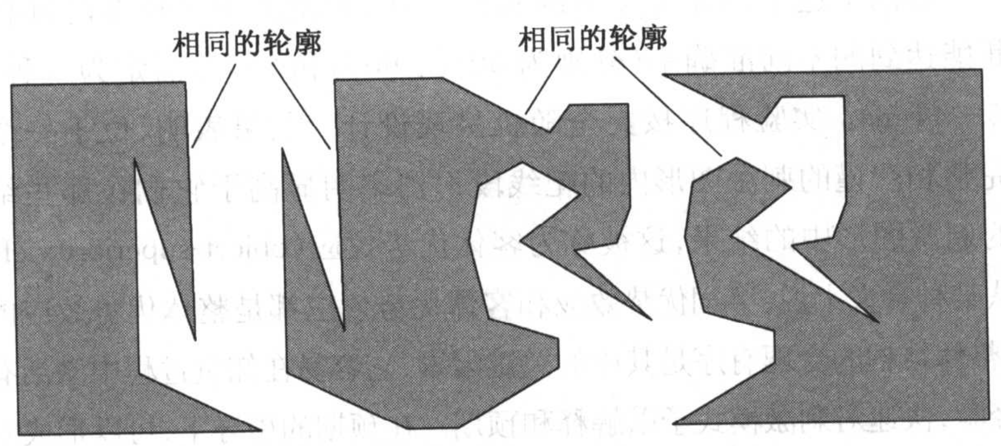


图 7-11　轮廓相同但形状可以不同

鲁宾把轮廓产生形状叫做"形状制造"(shape making),当视野被轮廓分为图形与背景时,轮廓只使图形有形状,而背景似乎没有形状。轮廓倾向于对它包围的面积发生影响,即一般是向内而不是向外发挥构成形状的效应。②

一般来说,轮廓都是出现在明度或色调发生突变的地方,但有些情况下,没有明度或色调突变的地方也会出现轮廓知觉,这就是主观轮廓(subjective outline)。主观轮廓的研究开始于 20 世纪初,1901 年舒曼(Schumann)第一次描述了主观轮廓的现象。1955 年卡尼查(Kanizsa, 1976)提出了主观轮廓的典型图示,即图 7-12。前文看到的主观轮廓图(图 7-2)是图 7-12 的较简化形式。从图 7-12 中,人们不仅可以看到三个黑色的圆盘、一个白色的三角形,还能看到一个有黑色边线的三角形。

对于主观轮廓的形成机制,多数心理学家用建构知觉理论来解释。在图

① 杨治良:《基础实验心理学》,甘肃人民出版社 1988 年版,第 223 页。

② 同上,第 224 页。

7-12提供的一系列刺激部件的作用下，人脑内部储存的相关知觉图式被激活，如圆盘、三角形，以及它们合理的遮挡关系。这些知觉图式可以为刺激模式提供最合理、最完美的解释，即使相关的信息并不完整。为了使构建起来的知觉印象突出、鲜明，观察者在主观上进行了信息的“补充”，这就出现了本来并不真实存在的“白色三角形的轮廓线”，这“轮廓线”还能从知觉印象中凸显出来。可见，知觉过程是知觉者在过去已有知识经验的作用下对当前刺激信息进行组织加工，从而建构“知觉印象”的过程，这一过程中存在着假设、检验和信息补充的心理过程。

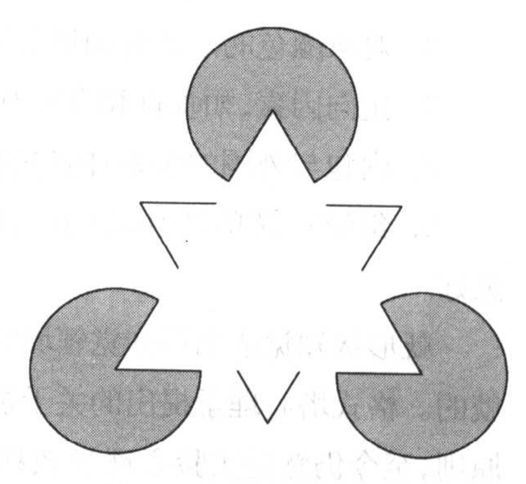


图 7-12 在无明度突变的地方形成了主观轮廓

此外，需要指出的是，马赫带效应在轮廓知觉形成中具有非常明显的作用，它使视野中明度或色调突变的地方更加显得界限分明。

(二) 形状

形状(shape)，就是轮廓从视野中分割出来的一个面积，也叫图形(figure)，视野中图形以外的区域叫背景(ground)。图形是视知觉的重要对象，也是人们认识外界物体空间特性的重要元素。最初对图形与背景进行区分的是鲁宾(Rubin, E. J.)，他的研究构成了形状知觉领域研究的基础。鲁宾确定了图形与背景的一系列差异性：①

1. 图形有形状，而背景相对地没有形状；
2. 背景似乎是在图形的后面连续伸展而未被中断；
3. 图形大都有某些特定物体的形状的特征，而背景好像是未成形的原料；
4. 图形倾向于在前面出现并凸出于背景之上，而背景好像是在后面，距离观察者较远；
5. 图形相对来说更为动人、好记和倾向于具有某种意义，而背景比较模糊，印象不鲜明。

格式塔心理学的代表人物之一考夫卡(Kurt Koffka, 1886—1941)进一步发展了鲁宾的学说，补充了一些新的图形与背景的差异性：

① 杨治良：《基础实验心理学》，甘肃人民出版社1988年版，第218—219页。

6. 对象颜色的刚柔对图形来说要比背景更重要；

7. 定向因素，如垂直和水平方向更容易组成图形；

8. 内包与外围的关系：内包的部分容易成为图形；

9. 组织的简单性和均匀性：图形与背景的配置造成的形状愈简单效果愈好。

就形状知觉或图形知觉领域的研究来说，格式塔心理学的工作是最富有成就的。格式塔心理学提出的关于视野中各种相互分离元素组织成图形的规律和原则，至今仍然是实验心理学教科书必须提及的内容。1923 年，格式塔心理学的代表人物韦特海默就开始用点子图研究分离的诸多点子如何组织成图形的问题，并试图揭示影响图形与背景形成的各种因素。综合许多学者后续的研究，已经发现的图形决定因素可概括为以下四个主要方面。

1. 空间接近性

同类物体在空间上较为接近时容易被知觉为一个整体，如图 7-13 所示。我们在看图 7-13(a)时，很容易看到一行一行的字母 A，因为字母排列的行间距大于列间距，所以一般会把距离较近的同一行的字母看成一个整体单元；在看图 7-13(b)时，容易看到一列一列的字母，因为字母排列的列间距明显大于行间距，所以一般会把距离较近的同一列字母看作是一个整体单元；在看图 7-13(c)时，由于字母的行间距和列间距比较接近，行知觉和列知觉印象的优势偏向不明显。

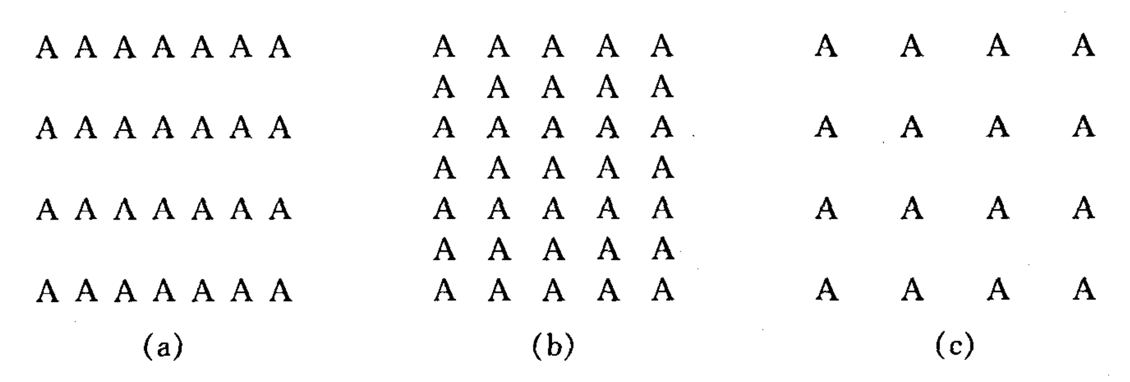


图 7-13 图形知觉中的空间接近性(spatial proximity)原则示意图

2. 相同或相似性

在亮度、色调和形状方面相同或相似的点子或图形倾向于结合在一起，形成一个较大的图形或单元，如图 7-14 所示。图 7-14(a)容易被看作由小圆圈构成的“十”字结构，在“十”字结构的每个象限中有一组黑色长方条；图 7-14(b)容易被看成一个由黑色点子构成的倒立三角形和一个由三个小圆圈构成的正立三角

形。相同形状的元素被组织成一个整体单元。

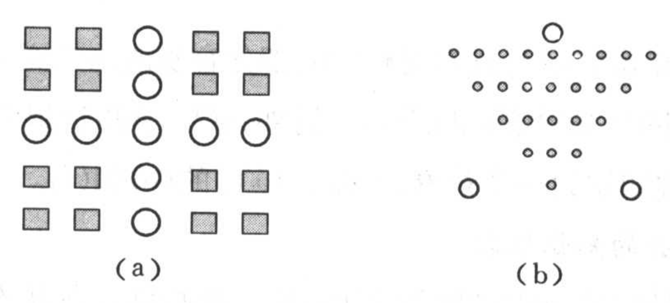


图 7-14 图形知觉中相同或相似性(similarity)原则示意图

3. 图形良好性

在知觉过程中,人总是试图利用已有的知识经验对当前刺激作出合理的、完美的、富于意义的解释,所以在对视野中的元素进行组织加工时,总是试图建构起合理的、完美的和富于意义的图形,这样的图形就叫做良好图形(good figure)。良好图形一般具有封闭、连续、对称等特征,如图 7-15 所示。

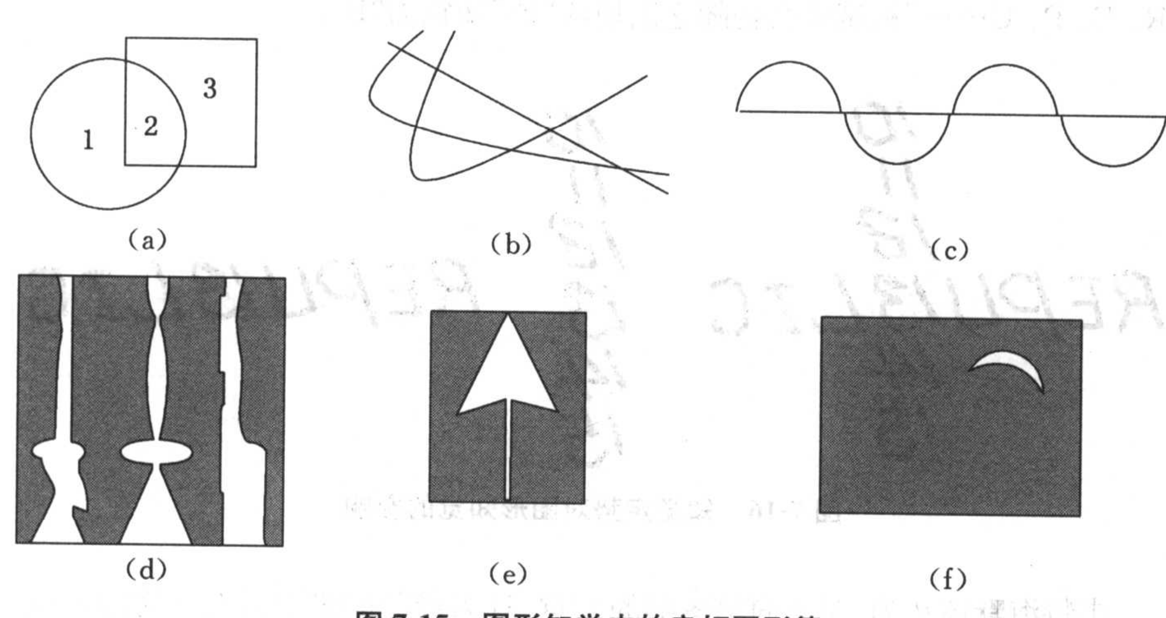


图 7-15 图形知觉中的良好图形律

封闭律(law of closure):有封闭轮廓的面积容易被知觉为一个完整的图形,如图 7-15 中的(a),人们一般会知觉到一个圆形和一个方形,而较少将其看成是 1、2、3 的三个不规则图形。只有知觉为圆形和方形才会得到封闭而又合理、完美的知觉结果。

连续律(law of continuity):具有连续的、方向性的部分容易被知觉为一个整体,如图 7-15 中的(b)和(c)。(b)容易被知觉为两条曲线和一条直线相互交叉,

而不是知觉为许多小的线段或小的面积;(c)则容易被知觉为一个连续的正弦波形。

对称律(law of balance):具有对称性的部分容易成为知觉的对象,非对称的部分成为背景,如图 7-15 中的(d)所示。另外,在图形知觉组织中,具有良好意义的部分也容易被知觉为一个对象,如图 7-15 中的(e)和(f)。

4. 定势与过去的知识经验

定势(set)就是在某一特定知觉活动中的一种特殊准备状态,这种状态可以在刚刚发生过的知觉过程影响下发展起来。定势作用与前述的各方面刺激因素有所不同,它是依个人主观条件而变化的因素,所以也可称为图形知觉组织的"非刺激因素"。[①]如被试从图 7-16 中知觉到的"13"或"B"能充分地说明这种定势的作用机制。当一个接一个地给被试呈现刺激,就可以建立起某种知觉定势,然后这种定势就会影响对后续刺激的知觉。如果连续给被试呈现"10、11、12……"被试就会把随之出现的"13"知觉为一个两位数;如果连续给被试呈现"R、E、P、U……"被试就会把随之出现的"13"知觉为"B"。

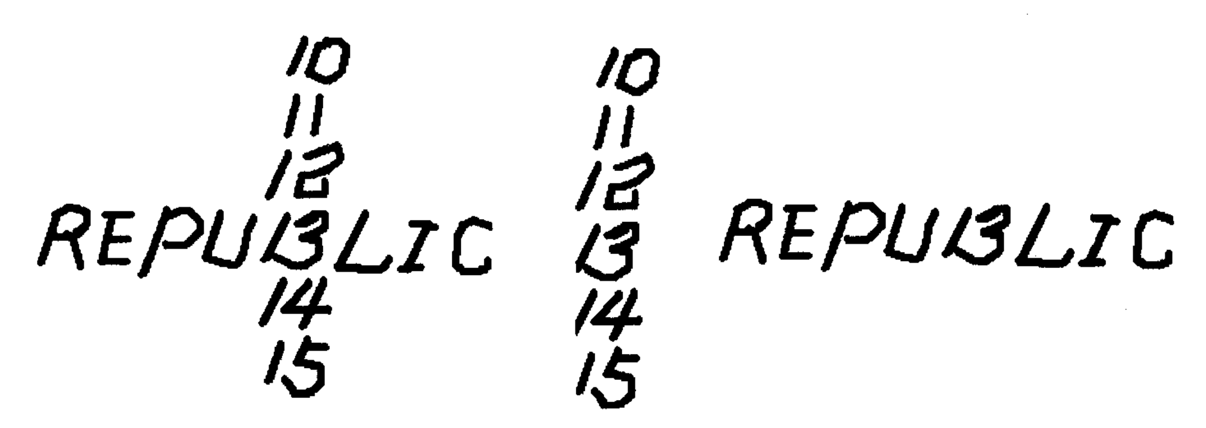


图 7-16 知觉定势对图形知觉的影响

韦特海默还认为,对于简单图形的知觉,过去经验的影响是有一定限度的,只有在它与刺激因素不存在矛盾时才起作用,如果二者发生冲突,一般仍然是刺激因素对知觉结果起到最为关键性的作用,也就是说,倾向于以刺激因素本身具有的特征而将其知觉为某一"良好图形",如图 7-17 所示。图 7-17 中的(a)的刺激因素掩蔽了"上",知识经验也难以使被试知觉到"上"的存在;同样,被试也难以从图 7-17(c)中知觉到"下",因为在这两个图形中"上"和"下"都参与到构建

① 杨治良:《基础实验心理学》,甘肃人民出版社 1988 年版,第 218—219 页。

"良好图形"的过程中。但从图 7-17 的(b)和(d)中就可以很容易地知觉到"上"和"下"字,因为此时知识经验能帮助被试克服多余刺激线条的干扰。

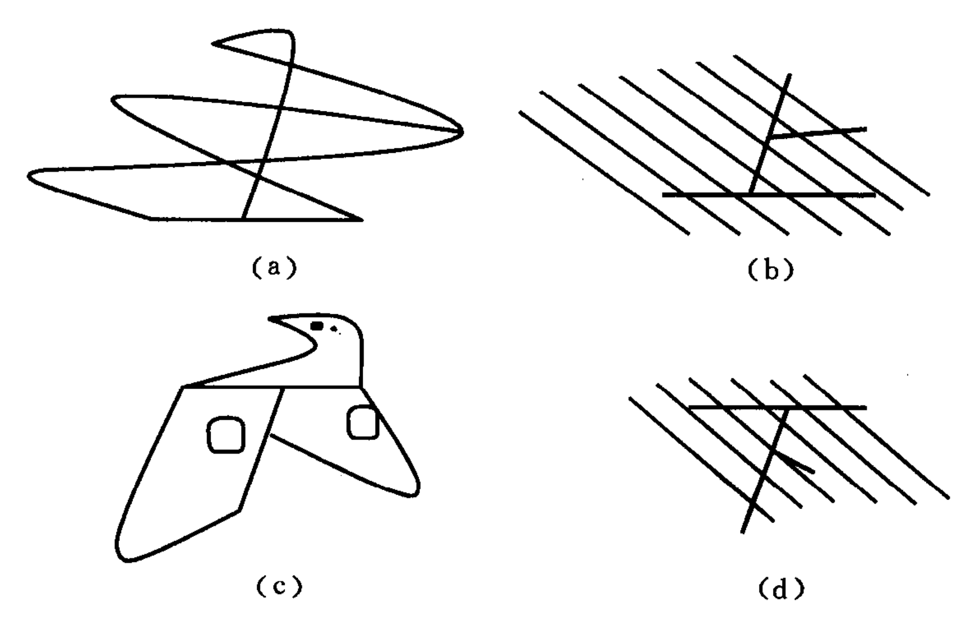


图 7-17　图形知觉中刺激因素与知识经验的竞争

二、图形掩蔽与后效[①]

知觉是需要时间的,因此刺激物之间的时间关联将会影响到对刺激物的视觉效果。我们将通过以下几种不同的现象来说明视知觉过程中的时间效应(temporal affter-effect)。

(一) 图形掩蔽

只要刺激在时间和空间上接近,那么刺激间就会相互干扰或相互掩蔽(masking),造成知觉彼此发生变异。在视觉掩蔽中,对目标刺激物的知觉会受到同时或继时出现的掩蔽刺激的影响而变得模糊或发生变异。根据掩蔽刺激呈现的时间不同,可将其划分为前项掩蔽(forward masking)和后项掩蔽(backward masking)。在目标刺激物出现之前的较短时间内,呈现掩蔽刺激物,从而对目标刺激物的知觉造成干扰,此种掩蔽就叫做前项掩蔽;在目标刺激物出现之后的较短时间内,呈现掩蔽刺激物,从而对目标刺激物的知觉造成干扰,此种掩蔽就叫做后项掩蔽。

① 朱滢:《实验心理学》,北京大学出版社 2000 年版,第 238—242 页。

1935年，沃纳(Werner, H.)在研究图形知觉时发现了一种有趣的现象，而这一现象对于研究轮廓的知觉过程和轮廓在成形中所起的作用是一种有力的手段。沃纳在视野的同一位置连续、迅速地呈现两个图形，一个黑方形和一个同样大小由黑框环绕的白方形，如图7-18所示。两个图形呈现的方式是，先呈现黑方形20 ms，经过150 ms的时间间歇(interstimulus interval, ISI)，时间间歇中只出现灰色背景，再呈现有框白方形，结果奇怪的现象发生了：被试竟然没有看到先呈现的黑方形。如果把两个图形呈现的顺序颠倒，则两个方形都能被看到；如果把两个图形的黑白颜色对调，也可以观察到同样的现象。可见，有框方形可以消除无框方形知觉，为什么呢？

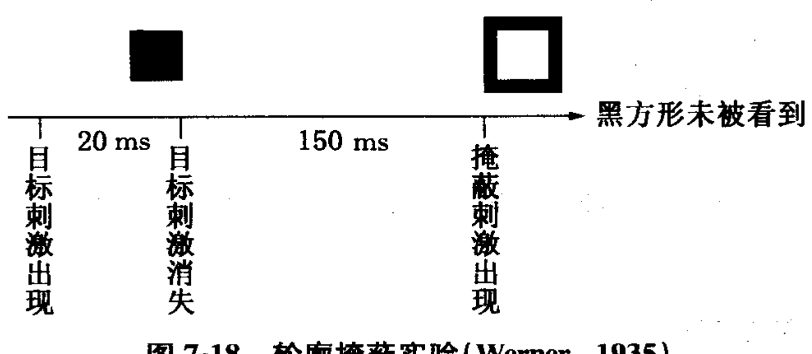


图7-18　轮廓掩蔽实验(Werner, 1935)

沃纳认为，当先呈现无框方形时，在无框方形还没有足够的时间建立起自己的轮廓之前，已被相反的有框方形的明度差别给抹杀了；但当有框方形先呈现，由于它的双重轮廓太强，所以就不致被消除掉。

进一步的研究发现，当目标刺激物与掩蔽刺激物在呈现时间上的间隔不同时，产生的掩蔽效果也不同：有时看不到目标刺激物，有时目标刺激物比较模糊，有时目标刺激物的结构不明显。产生上述效应的原因，朱滢等(2000)认为是视觉滞留(visual persistance)造成的。视觉滞留是指刺激作用后神经活动的惰性，这种神经活动的惰性使得刺激物在消失后，刺激物的作用痕迹依然存在一段时间。虽然短暂呈现的刺激物之间在时间和空间上相互分离，但是它们仍被知觉为同时出现，并相互干扰。

掩蔽也发生在目标刺激与掩蔽刺激同时呈现的情形中。加里和查马曾做过如下实验来考察图形掩蔽及其识别的过程。他们在纸上画出一个几何图形，这一几何图形可以是方形，也可以是圆形。然后在这个几何图形上叠加另一个或多个几何图形，以便在或大或小的程度上掩蔽第一图形，如图7-19所示。将这

个组合的图形呈现给被试，要被试找出其中隐含的目标图形，并描述这个过程，被试没有时间限制。从被试的表述中，加里等人了解到，被试先从组合的图形中找到一些可能暗含某一熟悉的或目标图形的线条，由这些线条构成一个简略图。如果这个简略图真的包含要寻找的部分，被掩蔽的图形就或多或少地被看出来啦。这种掩蔽实验中，目标刺激与掩蔽刺激是同时、持续存在的。

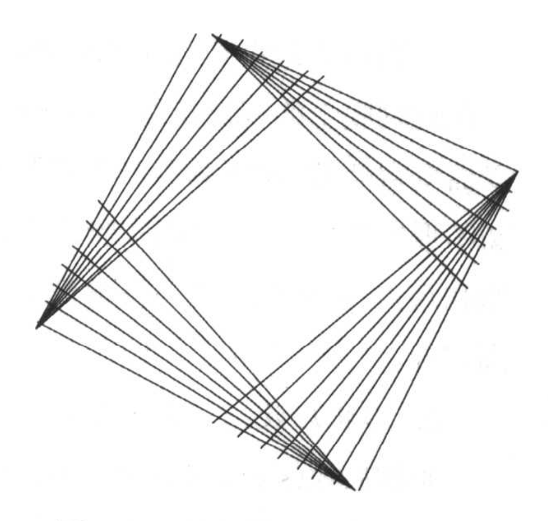


图 7-19　图形掩蔽(正方形被掩蔽)

这种图形掩蔽实验后经威特金(Witkin, H. A.)进一步研究和发展，形成了独特的认知方式"镶嵌图形测验"并沿用至今，对此在阅读材料中作专门介绍。

(二) 图形知觉中的后效

一种刺激作用之后对随后的心理活动的影响称为后效(after-effect)。后效有很多种，在视觉通道里表现的形式也是多种多样的。在此介绍图形后效、形状后效和关联后效。

1. 图形后效

图形后效(figural after-effect)是指注视一个图形一段时间后，对随后感知别的图形的影响。图形后效的经典例子是苛勒和沃莱奇(Köhler & Wallach)于1944年证实的，即一定形状的边界会从原来位置发生"位移"。如图 7-20 所示，如果被试先注视(a)图的×点 40 s，然后再把注视点转移到(b)图的×点，此时(b)图中左侧两个正方形之间的距离看起来要比右侧两个正方形之间的距离大，而实际上这两个距离是相等的。

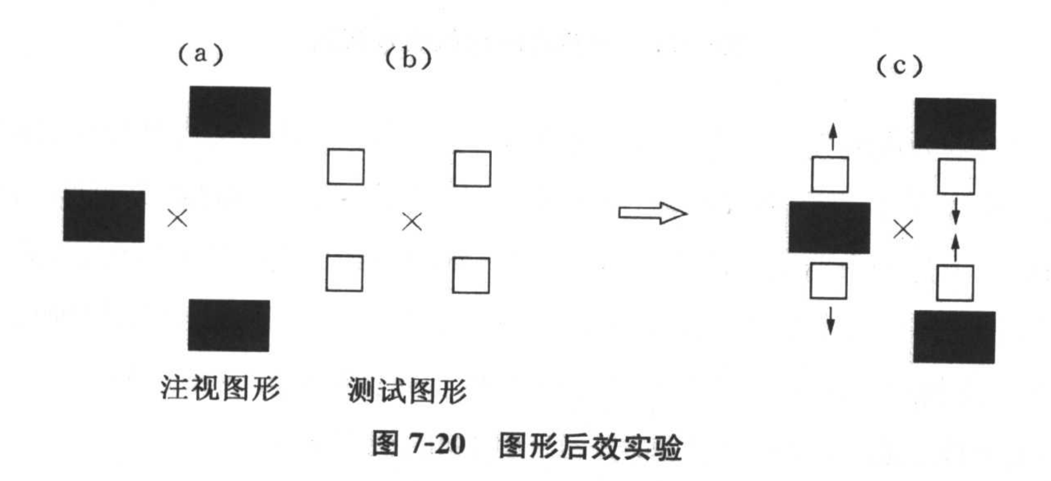


图 7-20　图形后效实验

苛勒和沃莱奇认为饱和(satiation)作用导致了这种位移,就是说,当一个图形在视网膜上的视像持续一段时间后,视网膜上的成像区域以及相邻区域的感受器变得疲劳,对新刺激产生抑制和饱和,使得投射到视网膜上饱和区域附近的新图形发生位移。(c)图实验模式表明,如果把(a)、(b)两图合在一起,那么根据饱和观点,左侧两个正方形分离、右边两个正方形聚拢,于是出现了知觉上的距离差异。

2. 形状后效——吉布森效应

形状后效(shape after-effect)又称为弯曲后效(curvature after-effect),是指由于图形方向导致的图形形状知觉的变化。吉布森(Gibson, J. J.)于 1923 年用实验证明了形状后效,所以该后效也被称为吉布森效应。吉布森让被试戴上一副特殊的棱镜眼镜,这种眼镜使进入的光线发生偏移,观察者会把垂直线段看成弯曲线段,如图 7-21 所示。观察者在注视弯曲线段的过程中,线段的弯曲度逐渐降低直至消失,即由棱镜眼镜产生的“弯曲”开始变得越来越直。然后摘下棱镜眼镜,就会因戴棱镜眼镜而产生后效,即会把直线看作是向棱镜眼镜产生的弯曲相反的方向弯曲。吉布森和雷德纳(Radner, M.)发现,在直接连续注视倾斜或弯曲线段一段时间后也会产生相同的后效:垂直线段向最初注视的倾斜或弯曲相反的方向倾斜或弯曲。

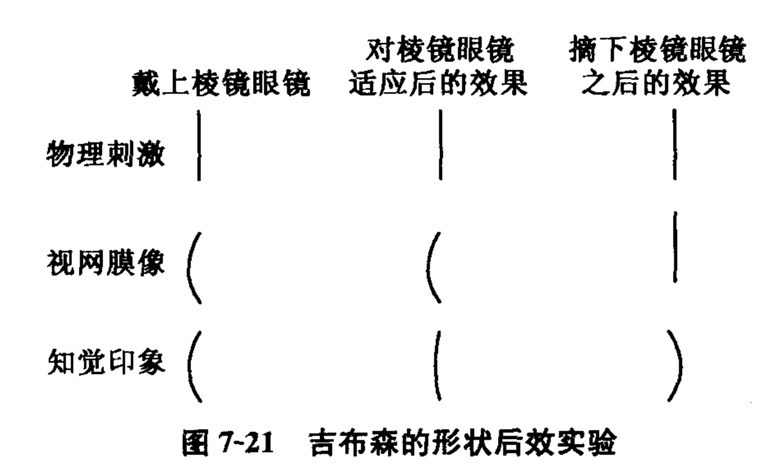
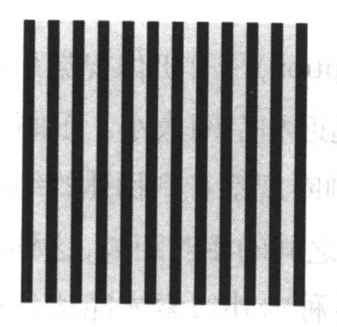


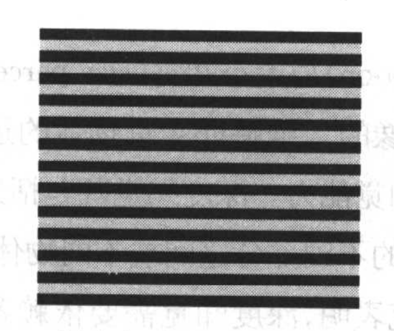
图 7-21　吉布森的形状后效实验

吉布森效应的发生是由于视觉系统对弯曲线段刺激的适应性反应,即当被试戴上棱镜眼镜之后,刺激线段在视网膜上形成了一个弯曲的视像,即视网膜接受的实际上是一个弯曲线段的刺激,当这一刺激持续作用较长时间时,视觉系统即发生适应性变化,弯曲的线段看上去不那么弯曲了。这时,如果摘掉棱镜眼镜,则线段刺激在网膜上形成不弯曲的视像,但是此时视觉系统还处于对弯曲刺激的适应性状态,于是产生了向另一方向弯曲的视觉映象。

3. 关联后效——麦科洛效应

关联效应(contingent after-effect)是不同刺激之间相互影响而发生的效应,如颜色和方向,或颜色和形状。1965 年,麦科洛(McCollough, C.)首先报告了这种后效,她给被试呈现几秒钟由黑色和橙色垂直栅条组成的图形,随后呈现黑色和蓝绿色的水平栅条组成的图形,两种图形交替呈现 4 分钟后,呈现测试图形。测试图形由垂直和水平的黑白栅条组成,如图 7-22 所示。观察者报告在测试图形中垂直方向黑白栅条处出现淡蓝绿色后效,在水平方向栅条处出现淡橙色后效。这种后效持续的时间可以长达几小时、几天甚至几个星期,而且在这段时间内被试可以自由观察其他事物,但这种后效并未因此消失。麦科洛使用的这两种颜色互为补色,当一种特定色调的颜色连续刺激视觉系统,视觉系统将对这种颜色发生适应,并产生其补色的感觉,即后效。这个实验的独特性在于此种后效是与图形轮廓的方向相关联的。


黑色与橙色垂直栅条


黑色与蓝绿水平栅条

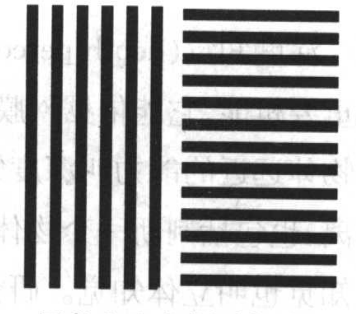

黑色和白色测试栅条

图 7-22　麦科洛效应

对麦科洛效应的进一步研究发现,如果把测试刺激栅条旋转 90 度,或者被试头部旋转 90 度,则后效颜色发生互换;如果测试刺激栅条或被试的头部旋转 45 度,则颜色后效消失。那么,这种关联效应是如何发生的呢?研究者认为,在视觉神经系统中存在各种各样的特征觉察器,这些特征觉察器各自独立地检测视野中刺激物的某种特征,如水平线特征、垂直线特征、角度特征、曲线特征,等等。在麦科洛的实验中,由于刺激材料中出现了"垂直线条"和"水平线条"两种重要特征,所以动用或激活了"垂直线条特征觉察器"和"水平线条特征觉察器",这两种特征觉察器各自独立地接受橙光刺激和蓝绿光刺激而发生适应,当换成白色背景后,它们分别对橙光和蓝绿光的感受性降低,于是主要接受到对应补色的作用,从而在知觉系统中出现原刺激色的补色效应。当刺激栅条或被试头部旋转 45 度时,对其知觉的特征觉察器是"斜线特征觉察器"。由于在原来的刺激

阶段,斜线特征觉察器未被激活或动用,所以也就不会对橙光和蓝绿光发生适应,自然不会出现颜色效应。看来,麦科洛的实验结果可以使用特征觉察器假说来加以解释,反过来,麦科洛效应也为特征觉察器假说提供了有力的证据。

麦科洛的这一发现扩大了视知觉研究领域,引发了一系列后续研究,比如有研究者给被试呈现两种都是水平的栅条,其中一个是绿色的并向上运动,另一个是红色的并向下运动,要求被试持续注视一段时间后使用黑白栅条进行测试。结果在测试阶段发现,黑白水平栅条向上运动时显红色,而同样的水平栅条向下运动时显绿色。这也是关联后效的一种,是随运动而产生的颜色反应。另外,使用扩张的和收缩的螺旋图作刺激,也可演示出随运动变化的彩色效应,也可出现颜色随运动变化的关联后效。①

第四节　深度知觉与大小知觉

深度知觉(depth perception)与大小知觉(size perception)是有机体视觉系统的重要机能,它能依据网膜像的二维特征知觉物体的远近和空间大小。眼睛判断物体远近的能力叫深度知觉能力。深度知觉既包括判断观察者到物体的绝对距离,也包括判断一个物体的不同部分之间或不同物体之间的相对距离,这种深度知觉也叫立体知觉。研究表明,深度知觉需要依赖各种内外线索才能达到相当的准确性。

一、深度知觉线索

人在空间视觉中可依靠多种客观条件和机体内部条件来判断物体的空间距离或位置,这些主客观条件统称为深度线索(depth cue),主要包括外界刺激的物理特征、单眼和双眼的生理机制、个体的知识经验等三方面。一般来说,深度知觉是在知识经验的基础上,整合了刺激因素和生理变化因素之后才形成的,其中知识经验起着极为关键的作用。个体在成长过程中,要不断地参与各种实践活动,在实践活动中尝试并校正距离或空间定位,随着年龄的增长、经验的积累,其深度知觉能力也不断发展。

---

① 张庆云:《普通实验心理学》,河南大学出版社 1993 年版,第 257 页。

根据在深度知觉中起作用的机制的不同,深度线索可以划分为两大类:非视觉性的生理线索和视觉性的深度线索。非视觉性的生理线索包括眼睛的调节和双眼视轴的辐合运动,这两个方面参照的是眼肌收缩或扩张的生理运动信息;视觉性的深度线索是那些经过机体加工、结构级差编码、整合后能成为判断空间位置依据的线索,又可分为单眼视觉深度线索和双眼视觉深度线索。单眼视觉线索主要包括刺激物的网膜像的大小、遮挡和重叠关系、线条透视、空气透视、光亮与阴影分布、颜色分布、运动视差等信息,双眼视觉线索主要是双眼视差。

(一)非视觉性的生理线索

1. 水晶体的调节运动

在注视远近不同的物体时,眼睛的睫状肌会自动改变其松紧度,调节水晶体的曲率以便透射光线能够聚焦在视网膜上,在这一过程中睫状肌及与其相连接的其他生理结构的生理运动信息传递到大脑皮层并得到加工,成为距离判断的线索。具体地说,在看远物时,睫状肌松弛,肌肉紧张度减小,水晶体变得比较扁平;在看近物时,睫状肌收缩,肌肉紧张度加强,使水晶体变得凸起。

大量的实验证明,眼睛的调节机能在深度知觉过程中的有效作用范围是10米以内,而且不太准确,因此这是一种相对较弱的深度线索。

2. 双眼视轴的辐合运动

双眼视轴的辐合运动(binocular convergence)就是指在看远近不同的物体时,双眼视轴的辐合角度发生相应变化,这种生理运动的信息传递到大脑皮层即可成为深度线索。具体地说,在看近距离物体时,双眼中央窝对准物体以保证来自物体的刺激光线聚焦于网膜的高感受区,双眼视轴趋向于聚拢,就形成较大的辐合角;在看远物时,来自刺激物的光线趋于平行,双眼视轴也要趋于分开以保证光线聚焦于视网膜上,视轴就形成较小的辐合角。可见,观察远近不同物体时,眼肌调节双眼以不同的角度进行辐合运动,这里眼肌的生理运动和眼球的辐合角变化信息传递到大脑皮层,遂成为深度判断依据。不过,双眼辐合运动在深度知觉过程中的有效作用范围大约在几十米以内。当刺激物过远时,双眼视轴近乎平行,刺激物距离的变化不能引起双眼辐合角的明显变化,它对估量距离就不起作用了。

(二)单眼视觉深度线索

单眼视觉深度线索主要是指物理刺激的空间分布特征,这些特征与个体经验结合就可使人们凭单眼视觉形成深度判断,所以这些刺激分布特征也叫做深

度知觉的物理线索(physical cues),主要包括:可形成的网膜像大小、物体的遮挡或重叠、线条透视、空气透视、光亮与阴影的分布、颜色分布、结构级差、运动视差等。

1. 可形成的网膜像大小

刺激物的实际大小是确定的,当其离开观察者的眼睛越远时,其在视网膜上成像就越小;当其离开观察者的距离越近时,其在视网膜上成像就越大。对于熟知的物体,在已经知道其实际大小的情况下,人们就可以根据其在视网膜上成像的大小来判断其远近。

按照视角原理,物体成像的视角与物体到观察者眼睛的距离成反比:

$$视角\ \alpha = (物体大小\ S) / 物体到眼睛距离\ D$$

当观察者与刺激物体的距离改变时,该物体在视网膜上成像的大小也会随之改变。由于人们在长期的生活实践中早已认识到了这一点,并将这种制约关系有意或无意地转换成一种大小知觉的自动编码机制,在物体距离、网像大小都发生改变时其对物体的知觉大小却保持不变,这在知觉心理学中被称为“大小一距离不变假设”。[①]当然,如果对被观察物体不熟悉,不知道其实际大小,也就难以依靠网像大小来知觉其远近了。

2. 遮挡或重叠

如果一个物体部分地遮挡了另一个物体,那么遮挡物会被知觉得近些,被遮挡物会被知觉得远些。如图 7-23 所示,圆锥体遮挡了长方体的一部分,观察者会很容易知觉到二者的远近关系:长方体离观察者更远些,这种深度线索在绘画中叫重叠(superposition)。根据遮挡或重叠关系,依靠单眼知觉物体的远近一般不会发生错误,但是这种线索难以帮助观察者判断物体之间的相对距离。

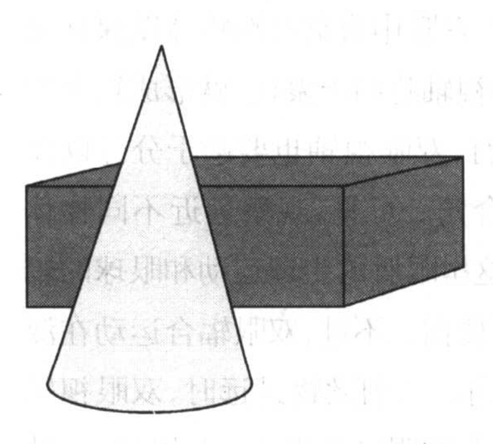


图 7-23 物体的遮挡或重叠

3. 线条透视和结构级差

线条透视(linear perspective)是指平面上的线条有汇聚趋向时,我们会把汇聚的一端看成是远离观察者的,而相互分离的

① 杨治良:《基础实验心理学》,甘肃人民出版社 1988 年版,第 256 页。

一端看成是逼近观察者的。这种情况多半出现在我们已知线条所表征的实际线条分布本应是平行或接近于平行的情况,典型例子就如图 7-24 中(a)所示的“钢轨”的空间分布等。结构级差(texture gradients)主要是指结构尺度大小在空间分布上按照梯级变化,观察者就容易将结构尺度较小的一端看成是较远的,而把结构尺度较大的一端看成是较近的,典型的例子就如图 7-24 中(b)所示的“地板”平面。

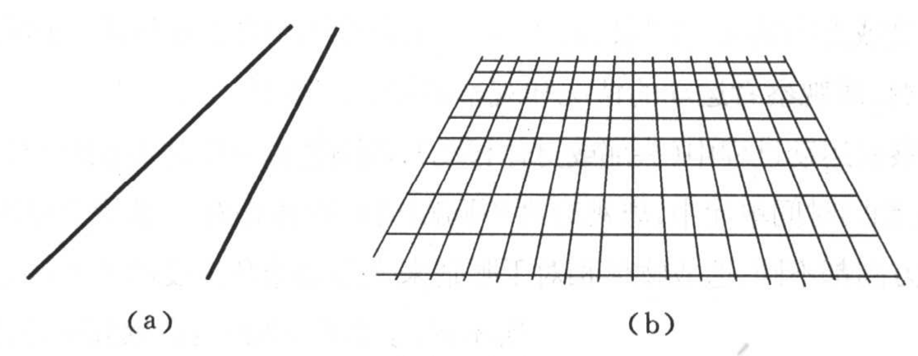


图 7-24　线条透视与结构级差

4. 空气透视

空气透视(aerial perspective)是由于空气中的灰尘和水分,使我们看到远处物体的轮廓较为模糊,近处物体的轮廓较为清晰,因此物体轮廓的清晰度成为深度知觉的线索之一。当我们看到物体的轮廓比较模糊时,就会觉得它远些;当我们看到物体的轮廓比较清晰时,就会觉得它近些。但是,根据清晰度来判断远近容易发生错误,因为空气中的灰尘和水分含量起伏变化较大,相当不稳定。比如,雨过天晴的时候,我们会明显感到以往轮廓朦胧的远山突然变得很近、轮廓鲜明,其实山还是那座山,它的距离还是那么远。

5. 光亮与阴影的分布

光亮的物体显得近,灰暗或阴影(shading)中的物体显得远,如图 7-25 所示。在绘画中,画家常常把近处的部分画得明亮些,而远处的部分画得灰暗些,以提高其三维立体效果。

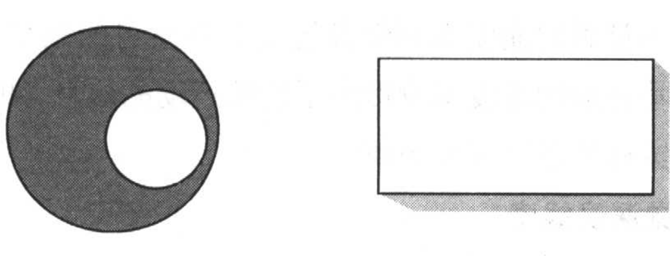


图 7-25　光亮与阴影

6. 颜色分布

在日常生活经验中，远方的物体一般呈现蓝色、近处物体呈现黄色或红色，所以我们通常会把蓝色与远距离、黄色或红色与近距离联系在一起，这也是画家在颜色使用中特别重视的一点。

7. 运动视差

运动视差(motion parallax)是指在运动过程中形成的对远近不同物体的不同运动速度或方向的知觉印象，它可以作为深度判断的重要线索。运动视差包括两种情况，即观察者运动的情况和观察者静止的情况。

在观察者运动过程中(头部或身体移动)，周围物体与眼睛的相对位置发生变化，而且在相同时间单元中，离观察者不同距离的物体相对于眼睛节点来说移动的角度和方向都不同，造成物体相对于眼睛来说运动的角速度的大小和方向都有所不同，于是产生视差：会感到比注视区更近的物体快速后移、注视区的物体保持相对静止、比注视区更远的物体则随同观察者一起向前，而且是越远的物体向前的速度越大。为什么会产生这种知觉差异呢？如图 7-26 所示，在某一小段的时间单元内，观察者乘车沿着公路从左向右运动，其在这一段时间内的注视区是 F 点，因此 F 点对于观察者的视轴来说，相对移动角度为 0，角速度也就为 0，观察者就会感到 F 区的物体保持相对静止。在比注视区更近的 Q 区，其相对于眼睛的视轴来说，移动角度不等于 0 且方向向后，观察者就会感到该区物体向后运动。相反，比注视区更远的 P 区，相对于眼睛的视轴来说，移动的角度不等于 0 且方向向前，观察者就会感到该区物体向前运动。

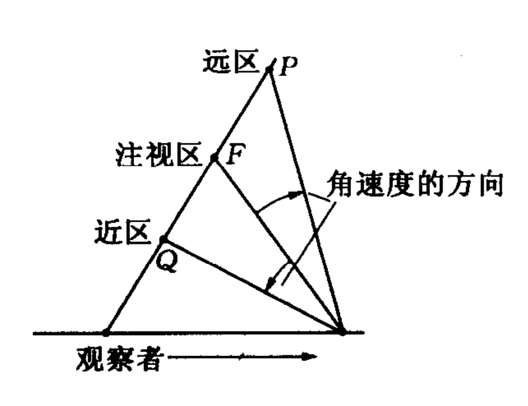


图 7-26　运动视差示意图

在观察者相对静止而物体运动的过程中，知觉到的速度也可以成为深度知觉线索。一般来说，以同样速度运动的物体，离观察者越近，知觉到的速度就越大；离观察者越远，知觉到的速度就越小，所以当我们熟知某物体运动的真实速度时，就可以依据知觉到的速度来判断其远近。观察者的眼睛追随远处物体的运动时，头部和眼睛转动的速度本身提供了物体运动的信息，知觉到物体运动的速度，进而依据知觉速度进行深度判断。

(三) 双眼视觉深度线索

深度知觉线索是多方面的，其中双眼视觉线索是基本的、重要的。通常，在

双眼视觉中,每只眼睛的视网膜上分别形成一个独立的视像,它们经神经信息传导到达中枢神经系统,并在中枢神经系统以特定的方式整合,产生一个单一的具有深度感的视觉映象。双眼视觉线索主要是指双眼视差,所谓双眼视差就是刺激物在双眼形成的视像差,它包括两种情况:一是同一刺激物因对于双眼来说存在视角差异而形成了不同的网膜像;二是物体在双眼网膜上成像的位置有所不同而形成双像。为了更好地理解双眼视差的这两种情况,我们需要首先介绍下述几个概念:视野单像区、双像、潘弄范围等。

1. **视野单像区**

研究表明,如果某一刺激物所在的位置点与两个眼睛节点连线的延长线交于两网膜的对应点上,则该物体形成单像。满足这一条件的各刺激点在空间形成一个通过两节点和注视点的圆周,该圆周被称为视野单像区(horopter)。如图 7-27 所示,*Q* 点是双眼的注视点,它的反射光线投射到两个眼睛的中央窝区域,形成清晰的单像;*P* 点和 *R* 点都处在单像区圆周上,它们的反射光线投射到两个眼睛视网膜的对应点上,也形成单像,但由于它们离开注视点较远,所以形成的单像较为模糊。*T* 点和 *S* 点则不在视野单像区圆周上,它们在两个眼睛视网膜上投射的位置都不对应,于是都形成双像或叫复视(diplopia),即 *T* 点和 *S* 点在双眼视野中都出现了两个视像。不过 *T* 点和 *S* 点形成的双像又有所不同:前者形成交叉双像,后者形成不交叉双像,对此,我们将结合图 7-28 来加以说明。

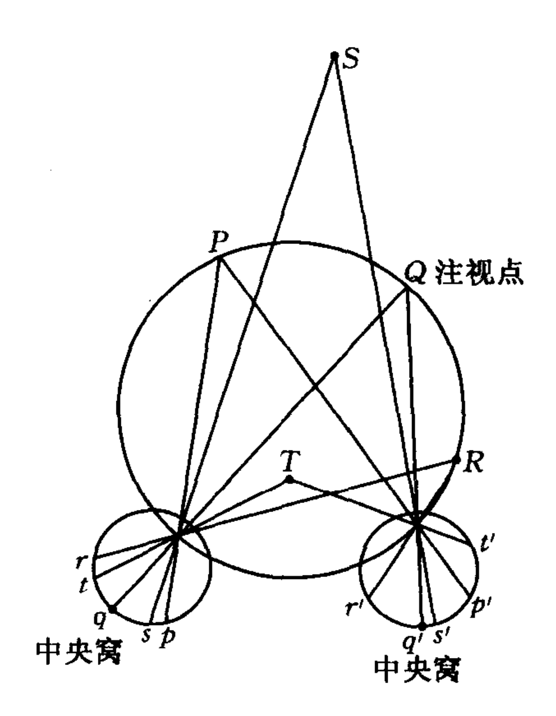


图 7-27 视野单像区圆周

2. **双像**

在某一时刻,眼睛的注视区相对稳定,而这时视野中其他各点成像情况相差很大。处在视野单像区圆周上的各点形成单像,而不在视野单像区上的各点形成双像,但形成双像的情况又各不相同。如图 7-28 中,有三种情况供比较:第一种情况是,当双眼注视 *A* 点时,与其相处在一个平面上且较近的 *C* 点和 *D* 点,在双眼视网膜成像的位置虽然不精确对应,但却差异不大,所以此时 *C* 点和 *D* 点

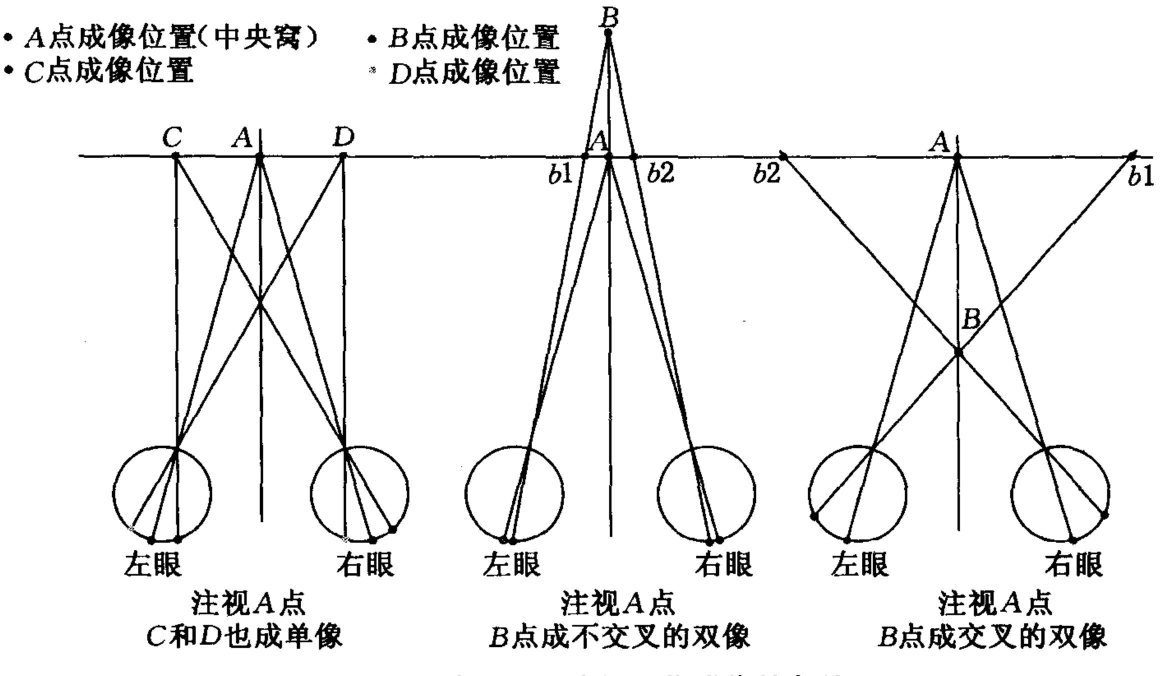


图 7-28　交叉和不交叉双像成像的条件

也形成单像;第二种情况是,当双眼注视 A 点时,比 A 更远的一点 B 在双眼视网膜上成像的位置相差较大,其成像在一个眼睛视网膜中央窝的左边,成像在另一个视网膜中央窝的右边,这时形成了不交叉的双像,即右眼成像在右边、左眼成像在左边;第三种情况是,当双眼注视 A 点时,比 A 更近的一点 B 在双眼视网膜上成像的位置也相差较大,其成像在一个眼睛视网膜中央窝的左边,成像在另一个视网膜中央窝的右边,这时形成了交叉双像,即右眼成像在左边、左眼成像在右边。

概括来看,当外界刺激模式不同的时候,其在视网膜上的成像模式也有所不同,这些双眼视差信息传递到大脑皮层,经过高级中枢的编码加工,就可形成立体知觉印象。

3. 潘弄范围

在前述的视野单像区与双像之间似乎存在某种矛盾,因为我们注意到:有时刺激物不处于视野单像区,但却也能形成单像。比如图 7-28 中的 C、A、D 三点处于一个平面,即使它们距离较近,但在注视 A 点时,C 和 D 点肯定不在通过 A 与两个眼睛节点三点决定的圆周上,可是它也形成了单像,我们把 C、A、D 三点知觉为在一个平面,如何解释这一矛盾呢? 1958 年,潘弄(Panum, P. L.)解决了这一矛盾,他认为当某一刺激在双眼网膜成像的位置虽不对应,但如果相互偏离未超出某一范围的限度时,仍可形成单像,这一范围就被称为“潘弄范围”或

"潘弄融合区"(Panum's fusional area)。在网膜的不同区域,潘弄范围的大小不同,越靠近中央窝的区域,潘弄范围越小,因此可以说中央窝区域对深度知觉更重要,它能更精细地反映刺激物的远近差异。

4. 立体物的双眼视差

两只眼睛观察平面物体时,这个物体在两个视网膜上的视像,基本落在相应部位上,而且在两个网膜上的视像模式基本一致,于是形成单像,产生平面物体的知觉印象。但是在观察一个立体对象时其知觉就不一样了,因为人的两只眼睛相距约 65 mm,这就使得两只眼睛会从不同的角度来观察这同一个物体,所以左眼看到立体物的左边多,右眼看到立体物的右边多,双眼会形成视差。两眼成像的这种差异,引起神经兴奋,并将这种兴奋传到大脑皮层,于是产生立体视觉(stereopsis)。一个对象或几个对象同时刺激双眼,是否形成双眼视差以及双眼视差的程度依下列条件而变化:第一,单一平面物体,或几个平面物体距离眼睛一样远时,双眼没有明显视差;第二,一个物体是立体的,或几个平面物离眼睛远近不同时,形成明显的双眼视差;第三,一个立体物的深度加大,或几个平面物的距离加大,双眼视差程度加大;第四,把立体物或不同的平面物,原来离眼睛远和近的部分对调,则左右眼网膜像也发生对调。

5. 深度视锐及其测量

双眼视差是深度判断的重要线索,但是人们利用双眼视觉来判断深度远近的能力也是有限度的,它存在一个差别阈限,即深度知觉的差别阈限。在心理学实验室内,可以利用深度知觉仪来测量人的深度知觉差别阈限。深度知觉仪(depth perception apparatus)及深度知觉差别阈限的测定如下所述。

最早测定深度知觉差别阈限使用的是赫尔姆霍兹设计的三针实验(three-needle experiment),后经仿造和改进,形成了霍瓦德-多尔曼知觉仪(Howard-Dolman perception apparatus),如图 7-29 所示。

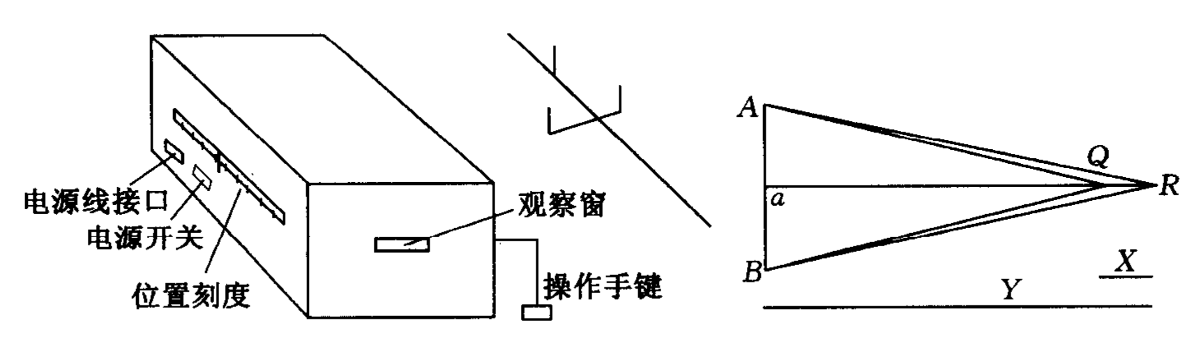


图 7-29 深度知觉仪及深度视锐的测定

霍瓦德-多尔曼知觉仪的基本结构：在一长方体的箱体内装一只日光灯管、三根立柱，在箱子的前端开一个观察窗，左边有一个刻度尺，如图 7-29 所示。三根立柱中，中间一根可以在水平轨道上滑行，两边的立柱固定在箱子内。当电源接通时，箱内的灯管点亮，被试坐在离观察窗一定的距离处，眼睛的高度与观察窗高度一致。实验时，被试通过观察窗判断中间的立柱与两边的立柱是否在一个平面上，即与观察者的距离是否一致。如果感到距离不一样，则可以通过操作手键来调节可调立柱的位置，以此方法可以测得被试深度知觉的差别阈限。假如测量得到的深度差别阈限为 X。很明显，X 越大，双眼视觉判断远近的能力越小；X 越小，双眼视觉判断远近的能力越大。此外，从图 7-29 还可看出，当 X 较大时，Q 点和 R 点形成的视角差越大，反之越小，因此我们可以使用这两个视角差作为深度判断能力的测定指标，而且把利用双眼视觉判断深度的能力叫做深度视锐(depth visual acuity)，深度视锐用前述的视角差来测量，二者成反比关系。如果用 n 代表这里的视角差，也叫像差角，则：

$$n=\frac{a}{(y-x)}-\frac{a}{y}=\frac{ax}{y(y-x)}(\text{弧度})$$

把弧度换算成秒的单位，则：

$$n=\frac{206\ 265ax}{y(y-x)}(\text{秒})$$

已有研究表明，人眼对距离的辨别灵敏度是很高的。虞积生等人对 1 159 名男青年的研究发现在距离为 6 米时，被试深度知觉差别阈限形成的视角差平均为 2.94 秒，标准差为 1.79 秒。[①②]也有人研究，对一个视力正常的人来说，经过训练可以辨别到 $n=2$ 秒的深度距离。

很显然，双眼深度视锐与照明条件有关，好的照明条件可以提高深度视锐。在较暗的照明条件下，深度辨别成绩较差，黑暗环境中双眼视差几乎不起作用。

此外，当观察者与物体的距离超出 1 300～1 500 米时，两眼的视线几乎平行，双眼视差近于零，这时双眼视差也就基本上不起作用了。但是在 1 500 米以

① 虞积生、方俐洛、张嘉棠：《中国正常男青年的深度视觉阈限值的测定(Ⅰ)》，《心理学报》，1980，12(3)：298—302。

② 方俐洛、虞积生、高晶：《中国正常男青年的深度视觉阈限值的测定(Ⅱ)》，《心理学报》，1980，12(3)：303—306。

外的物体,我们仍能知觉出远近,这是因为除了双眼视差的深度线索,还有其他的深度线索,如前文所述的视像大小、物体遮挡、颜色分布、空气透视等物理线索,即单眼视觉线索。

二、大小知觉与恒常性系数

日常生活中,人们能够比较准确地知觉不同距离上物体的实际大小,并且知觉大小既不完全按视角规律而定,也不按物体的实际大小而定,而是以介于这两种机制之间并偏向于恒常性一端的机制而定。假如在距观察者 1 米远的地方立一根 2 米高的标杆,观察者从距标杆 1 米处走到 5 米处,回头再看标杆。如果只按视角规律来计算,知觉到的标杆大小应缩小到原来知觉大小的 1/5,即 0.4 米,但被试知觉到的大小可能不是 0.4 米而是 1.6 米,甚至更高,接近于 2 米。看来,在一定的距离范围内,对刺激物的知觉大小基本上符合恒常性规律,而这种知觉恒常性(perception constancy)也可以通过测量人们对物体的知觉大小进行测量。

(一) 大小知觉恒常性系数的测定

可按照布伦斯维克比率(Brunswik Ratio,简称 BR)(Brunswik, 1929)或邵勒斯比率(Thouless Ratio,简称 TR)(Thouless, 1931)来测定和计算大小知觉的恒常性系数(constancy ratio)。其计算公式如下:

$$BR = \frac{(R-S)}{(C-S)}$$

式中,C 为刺激物体的实际大小、R 为被试对刺激物的知觉大小、S 为刺激物距离 1 米处的视角透视大小。比如,一标杆的实际高度是 1.5 米,一被试距标竿 5 米,要求其使用米尺复制标杆的高度,复制的结果为 1.3 米,那么该被试的知觉恒常性系数是多少呢?

如图 7-30 所示,物体在 1 米处的视角透视大小 $S = 1.5 \times (1/5) = 0.3$ 米,于是将上述数据代入布伦斯维克比率计算公式,得到的恒常性系数为 0.83。

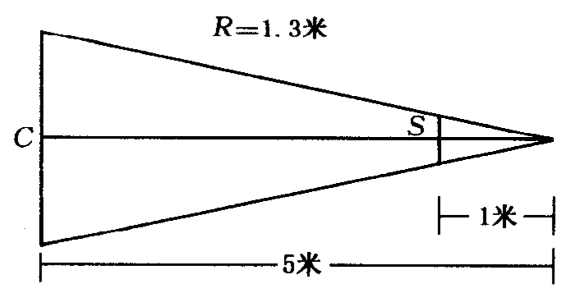


图 7-30 大小知觉恒常性系数的测定

一般来说,布伦斯维克恒常性系数的取值范围为 0~1。当 $BR = 0$ 时,说

明知觉中不存在任何的恒常性；当 $BR = 1$ 时，说明知觉中具有完全的恒常性。而一般我们实际计算的既不是 $BR = 0$，也不是 $BR = 1$，正说明人们不是按照视角规律知觉大小，也不是按照物体实际大小来完成知觉，而是介于二者之间，即布伦斯维克比率介于 0～1，且偏向于 $BR = 1$ 的一端。

邵勒斯比率与布伦斯维克比率表示的意义一样，不同点在于邵勒斯比率中使用 C、R、S 的对数进行计算，其取值范围也是 0～1，用 TR 表示。

$$TR = \frac{(\log R - \log S)}{(\log C - \log S)}$$

应用上述两个公式可研究大小、形状、颜色和亮度的恒常性，以及计算知觉中的恒常性系数。通常，在大小知觉中用得更多的是布伦斯维克比率，亮度知觉中用得更多的则是邵勒斯比率，因为在亮度知觉中，物理亮度与知觉亮度成对数关系。邵勒斯之所以修改布伦斯维克比率，就是为了使它更适合如知觉亮度与物理亮度的对数关系。

（二）经验与大小知觉恒常性

早在 1941 年，霍威和波林（Holway, A. N. & Boring, E. G.）进行了一项成功的实验研究，证实了经验对大小知觉恒常性的影响。他们的实验是在一个 L 形的走廊中完成的。[①]如图 7-31 所示，O 为被试的位置、S_C 为比较刺激、S_S 为标准刺激、D_C 为比较刺激与被试间的距离、D_S 为标准刺激与被试间的距离。比较刺激和标准刺激都是用幻灯在白幕上照亮的圆点。比较刺激距离被试 10 尺，在整个实验中距离保持不变，即 D_C 是恒定的，但为了与标准距离匹配，圆点的大

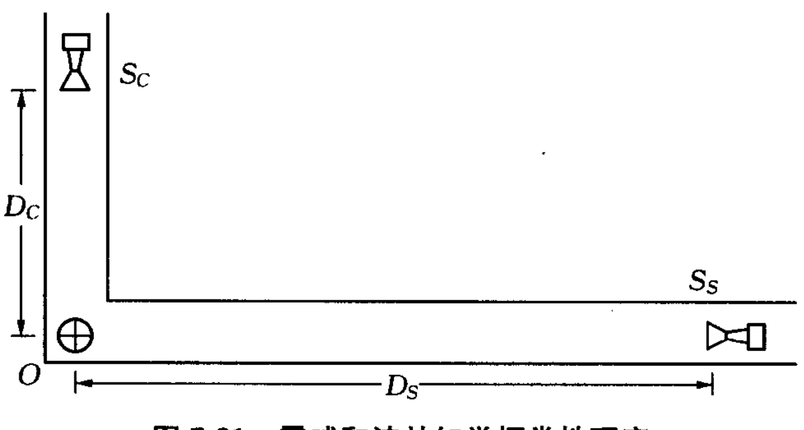


图 7-31　霍威和波林知觉恒常性研究

① 彭聃龄：《普通心理学》，北京师范大学出版社 2001 年版，第 143—144 页。

小可以随时改变。标准刺激与被试的距离从10尺到120尺不定，但是无论距离（Ds）怎样改变，作为标准刺激的圆点的大小与观察者的眼睛始终保持1分视角。另外，比较刺激和标准刺激强度，对被试的眼睛来说是相等的。实验进行时，主试以渐增法或渐减法改变比较刺激，并且根据被试的要求可以随意调整比较刺激，直到他认为比较刺激与标准刺激相等为止。

实验是在以下四种条件下进行的：

(1) 双眼观察：正常的双眼观察；

(2) 单眼观察：只准许使用单眼观察；

(3) 单眼用人工瞳孔(artifical pupil)观察：即用单眼通过一个小孔观察；

(4) 单眼用人工瞳孔通过减光筒观察：在遮挡视觉光路周围的光线和其他线索，相当于被试通过一个长长的遮光通道观察标准刺激物。

实验结果如图7-32所示。从图中1、2两条直线看，双眼和单眼观察条件下，都明显地保持着大小知觉的恒常性；从3、4两条直线看，这两种观察条件下，由于在相当程度上排除了周围环境的参考深度线索，知觉大小偏向于按视角规律而定，恒常性几乎不存在了。由此看来，在大小知觉中周围环境的线索起着关键性的作用，个体借助环境提供的深度线索来判断知觉对象的距离，在此基础上知觉它的大小。日常生活中，知觉对象的实际大小是熟悉的，如人、房屋等，或是在熟悉的环境中，估计对象的大小时可参考熟悉的物体，从而保持大小知觉的恒常性。

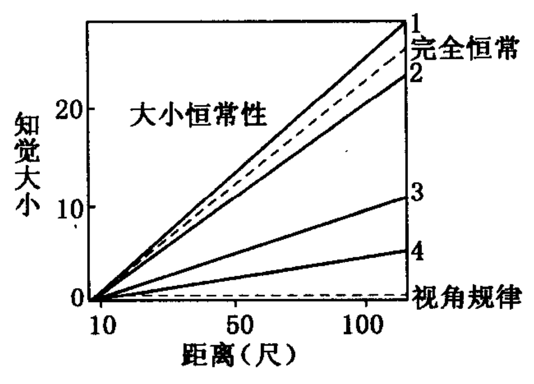


图7-32　霍威和波林的大小恒常性实验

1. 双眼视觉　2. 单眼视觉
3. 单眼使用人工瞳孔观察
4. 单眼使用人工瞳孔加减光筒观察

当知觉对象很远时，网像大小的作用逐渐增加，大小恒常性的作用逐渐减小，即知觉大小愈来愈符合光学的透视原理。比如把1千米的对象知觉为1，那么对象在2千米处就会被知觉为1/2、在1万米就会被知觉为1/10，同一个对象的知觉大小按照视角规律而变化。另外，水平观察和垂直观察时，大小知觉恒常性也不一样，水平观察时恒常性较大，垂直观察时恒常性较小，比如从楼顶向下观察物体比在同样距离下水平观察时要小得多。

阅读材料 7-1

汉字识别的实验研究

汉语是世界上使用人数最多的一种语言。作为世界语言文字大家庭中的一员，汉语汉字既有与其他语言文字相似性的一面，也有不同于其他语言文字的地方。因此，开展汉语和其他语言认知的比较研究，引起了国际上许多学者的重视。

20世纪80年代中期以来，我们以汉字识别为重点，在汉语认知方面开展了比较系统和深入的研究。在研究中我们遵循着下面一些原则：1. 信息加工的原则：我们把人脑加工汉语的过程看成是信息加工的过程。2. 层次原则：对汉语认知的研究是在四个层次上进行的，即复杂行为层次、信息加工层次、计算机模拟层次和神经生理学层次。3. 结构和过程统一的原则：通过反应时间记录技术和ERP技术研究汉语信息加工的过程，同时从反应时间和正确率的结果中推论人脑存储汉语信息的内部结构。4. 汇聚一致的原则：通过综合地运用反应时间记录技术、脑电分析技术、脑成像技术、计算机模拟技术在不同层面上研究汉语认知的内部结构和过程。5. 对比原则：在汉语和其他语言的对比研究中揭示汉语认知的特点和规律。

什么是汉字识别的基本单元？我们在一系列实验中，用词汇判断作业和命名作业，研究了在不同语境作用下汉字笔画数对汉字识别的影响，笔画和部件、语音和字形、词频和语境在汉字识别中的作用。这些研究汇聚一致地表明，在汉字识别中存在显著的笔画数效应和语境效应，词(字)频和语境也存在显著的交互作用。这说明在汉字识别中，存在特征分析的阶段；汉字识别有不同的层次，最低层是笔画层，其次是部件层，再次是字层；字的识别依赖视觉特征单元的激活，也依赖语境的作用。汉字识别是两种因素交互作用的结果。

汉字是一种表义文字或音义文字。在汉字中，形声字占90%以上。形声字用声旁表音，用形旁表义，这是汉字不同于其他文字的重要特点。在汉字识别中我们关心的另一个问题是形声字读音和语义提取的问题。

我们在一系列实验中，对比研究了不同频率条件下一致形声字、不一致形声字、例外字和非形声字的命名。结果表明，频率与字的类型有显著的交互作用。

这和前人的研究结果一致。同时,我们还发现了邻近字(敌对字与邻近字)频率的作用和声旁的独立性效应。根据以上结果,我们提出,汉字形声字的读音是由两种因素共同决定的,一个是声旁的读音,它提供了字下水平的语音线索,另一个是邻近字的读音,它提供了字水平的读音线索。一个目标字的读音是这两种因素综合起作用的结果。由于存在声旁的独立性效应,这一结果很难用并行分布加工模型(PDP 模型)进行解释。根据实验结果,我们提出了"高频优先"和"整字优先"的原则。"高频优先"是指,高频字的语音激活先于低频字;"整字优先"是指,整字的语音激活先于声旁的语音激活。我们还研究了形声字形旁的认知,发现形旁在字的范畴义的提取中有重要作用。这些结果对说明"亚字水平"信息的作用也有重要意义。

我们还研究了语音的自动激活及其在字词识别中的作用。在一项早期的研究中我们发现,"见形知音"和"知音识义"都需要花费一定的时间。由于音义之间没有显著差异,我们认为,音义可能同时发生,有音必有义,有义必有音,它们无法在时间进程上区分开来。随着阅读熟练的提高,知音识义的时间可能变得越来越短。

——资料来源:《汉字认知研究——15 年工作的回顾》(彭聃龄,2001)①

阅读材料 7-2

认知方式及其相关研究

认知方式(cognitive style),又称认知风格,"是人们对信息和经验进行组织加工时所表现出来的个别差异,是个人在感知、记忆和思维过程中经常采取的、习惯化的态度和风格"②,也是个体在组织和表征信息时所具有的偏好或习惯性方式。③自 20 世纪 40 年代开始,美国心理学家威特金(Witkin, H. A.)就空间知觉误差问题展开了一系列研究,连续设计了斜屋斜椅实验、转椅实验、棒框实验

① 彭聃龄:《汉语认知研究——15 年工作的回顾》,见中国心理学会:《当代中国心理学》,人民教育出版社 2001 年版,第 68—69 页。

② 张厚粲等:《关于认知方式的测验研究》,《心理科学通讯》,1982(2)。

③ Tennant, M. (1988). *Psychology and adult learning*. London: Rontledge, 100.

和镶嵌图形测验，并从中发现了个体在这些实验中知觉误差的一致性，于是提出了场依存性认知方式(field-dependent cognitive style，简称FDCS)的概念。随后的研究逐渐突破知觉领域而扩大到思维、想象、记忆、社会认知与社会交往，甚至学科偏向与职业定向等问题。显然，认知方式已被看作是对个体活动具有广泛影响的一个人格维度或人格特质。20世纪80年代前后，国内外学术界出现一个认知方式研究的热潮，进而引发了对认知方式的本质及认知方式测评方法的争论。

就连威特金本人也先后给出了若干不同的认知方式定义。最初，威特金给出的定义是：依赖外在视野还是依靠内部感觉经验的知觉倾向性。这一定义明显是缘于他在垂直知觉实验中的发现，但是不久，他得到垂直知觉误差与镶嵌图形测验分数间的高度相关，就将认知方式的定义修订为：分析的或整体的知觉去隐蔽能力。当他在知觉领域之外有了一系列重要发现之后，又将认知方式定义为：依靠外部问题情境的参照还是依靠内部自我体验的参照的倾向性。根据实验研究的结果，威特金及其同事确信人们在认知活动中是具有场依存性-独立性差异的，并假定这种个别差异可以用独立于周围环境影响而进行自主活动能力来加以说明。因此，尽管不同的心理学家研究所采取的角度和所得到的研究结论都不尽相同，但关于认知方式的特征基本能形成以下几点共识：

第一，过程性：认知方式是过程变量而不是内容变量，即指向认知过程而不是认知内容；

第二，普遍性：场依存性-独立性认知方式体现在各种认知操作中，影响是广泛的；

第三，稳定性：个体发展中，认知方式的相对水平是稳定的，比如威特金等人曾对1578名学生进行了长达10年的跟踪研究，结果表明绝大多数被试认知方式的相对水平无明显变化。

第四，中性：场依存性-独立性认知方式在价值上是中性的，无好坏之分，维度的任何一端都对环境的某些方面有适应价值。随着研究的逐渐丰富和深入，认知方式的概念有所发展，但这四个特征作为概念的核心被广泛认同。①

---

① Witkin, A. A., Moore, C. A. & Goodenough, D. R. (1977). Field dependent and field independent cognitive style and their educational implications. *Review of Educational Research*, 47: 1—64.

认知方式的稳定性是否意味着其内在的遗传性或受到生理基础的制约呢?对此也有了不少的研究。已有生理学、遗传学和神经科学的研究表明,激素水平、X染色体特征、脑机能单侧化、脑电活动都可能与认知方式有关。布罗弗门(Brovermen, 1964)和瓦伯(Waber, 1976)等人在激素水平上的研究发现,如果个体酮类固醇含量高,其认知改组能力会下降,而女性青春期及成熟较早,酮类固醇含量高,心理分化程度低,因而同一年龄的女性比男性更倾向于场依存性。也有研究者从染色体上的基因特征进行研究,发现与空间能力相关联的基因大部分存在于X染色体并随着X染色体的移动而传递。如斯塔福德(Stafford, 1961)和博克(Bock, 1973)的研究发现,母亲与儿子之间、父亲与女儿之间在空间知觉能力方面存在显著相关,而父亲与儿子间的相关几乎接近于零。很显然,这些相关关系与遗传的X染色体的基因控制假设是一致的。[①]古迪纳夫(Goodenough, 1977)等人的研究发现,具有相同X染色体的兄弟在棒框测验和镶嵌图形测验中的相关明显高于具有不同染色体的兄弟。这说明场依存性-独立性特质确与X染色体所携带的基因有关。

最近,关于认知方式的研究已不再囿于场依存性认知方式,而是从具有更广泛意义的"分析-综合维度"进行探索,这些探索还借助现代认知神经科学实验技术意欲察明认知方式的脑生理基础。科恩(Cohen)等人研究认为认知方式可能与脑机能的单侧化有关。他们的实验结果表明:大脑两半球对外界信息加工的方式和机制有所不同,左半球是序列性的分析风格,对信息进行继时的序列加工;右半球是平行性的综合风格,对信息进行同时的平行加工。而不同的人的左右半球优势对比存在差异,从而表现出不同的认知方式。科恩提出大脑左右半球上电活动的个体差异可能会造成个体认知方式的差异性。利维(Levy, 1990)甚至可以通过测量个体在休息状态时大脑半球的兴奋性激活的平衡,来预测其言语和空间行为表现。人在休息时脑电图中α波的不对称,可能与认知方式的言语-表象维度相对应。赖丁(Riding, R.)等人设计实验考察了不同认知方式的人在信息加工过程中大脑皮层15个不同部位的α波(可以表明心理活动的发生),结果表明:如果按照整体-分析维度来区分,分析型被试在所有任务中其15个部位的α波都低于整体型被试的;如果从言语-表象维度上来区分,言语型被

① 宋广文、李寿欣:《场依存-独立性认知方式:理论演进及其应用研究》,《内蒙古师范大学学报》,1999(2)。

试在大脑左半球有相对更高的α波抑制，而表象型被试恰恰相反，其大脑右半球α波抑制更高，存在明显的单侧化效应。总之，通过脑电记录与分析，我们可以确定认知方式是存在脑生理基础的，但是其中的对应关系尚不明朗，还需要进一步的精细研究。①

近来，研究者对认知方式的关注程度又有提升，引发了不少有价值的研究课题，同时也使得究竟如何对认知方式进行测量成为一个关键问题。国内学者更多采用的方法有棒框测验（rod-frame test）②和镶嵌图形测验（embedded figure test）③。但是，越来越多的学者认为，这两类测验都不能反映认知方式的全部内容，况且这两类测验本身也都存在技术问题且数据都存在较大的误差，就连这两类测验之间的相关系数也并不高。有研究得到0.49的相关系数，虽然达到了统计学上的显著性，但依然不是太高的效标效度，这意味着什么呢？是其中一个测验的效度较低，还是两个测验的效度都较低呢？仍需要系统研究。目前来看，赖丁的研究及提出的测量方法值得关注，他在认知控制模型里把认知方式置于一种认知控制水平，认为认知方式在个体内部世界和外部状态交互作用的界面上起着重要的组织和表征作用。赖丁从两个维度上分析和测量认知方式：整体-分析维度、言语-表象维度，前者与信息的组织过程有关，后者与信息的表征方式有关。

——资料来源：曾晓尤④，邓铸（2005）

建议阅读文献

1. 郭秀艳：《实验心理学》，人民教育出版社2004年版，第355—430页。
2. 余嘉元、邓铸：《当代认知心理学》，江苏教育出版社2001年版，第97—125页。
3. R. J. Sternberg著，杨柄钧等译：《认知心理学》，中国轻工业出版社2006年版，第83—110页。

① R. 赖丁、S. 雷纳著，庞维国译：《认知风格与学习策略》，华东师范大学出版社2003年版。
② 杨治良：《实验心理学》，浙江教育出版社1998年版，第598—600页。
③ 杨博民：《心理实验纲要》，北京大学出版社1989年版，第384—388页。
④ 曾晓尤：南京师范大学教科院2004级在读硕士研究生，方向为“发展与教育心理学”。

复习思考题

1. 如何理解形状知觉、轮廓、主观轮廓、知觉定势、图形掩蔽、吉布森效应、麦科洛效应、深度知觉、双眼线索、视野单像区、潘弄范围、双眼视差、深度视锐、时间知觉、运动知觉？
2. 简述知觉的直通理论、建构理论、计算理论。
3. 人类知觉加工的主要方式有哪些？有哪些主要特征？
4. 格式塔心理学提出的关于形状形成的原则有哪些？
5. 主观轮廓形成的心理机制是什么？
6. 如何解释吉布森效应和麦科洛效应的形成机制？
7. 深度知觉的线索有哪些？
8. 双眼视差是如何形成的？
9. 深度视锐是如何测量和计算的？
10. 主要的单眼视觉深度线索有哪些？
11. 大小与距离的关系如何？如何测量和计算大小知觉恒常性系数？

第八章
记忆的实验研究

本章内容提要

记忆是过去经验在人脑中的反映,是人类学习与工作的重要条件之一,它能将人类心理生活的过去、现在和未来连接起来构成整体,使人的认知及人格品质得以在其中形成和发展。人通过感知活动把外部信息输入记忆系统进行编码、存储和提取,借此积累知识、保存经验并加以应用。本章介绍记忆实验研究的主要方法,具体包括:(1)记忆的实验研究简史,介绍艾宾浩斯的开创性工作,经典记忆实验研究范式的发展,以及现代认知心理学对其研究内容、研究手段与研究方法的不断拓展;(2)记忆的基本实验方法,包括记忆实验材料的选择、呈现以及保持量的检测等;(3)记忆结构的经典实验研究,介绍短时记忆、感觉记忆的发现及相关理论;(4)内隐记忆的实验研究,介绍内隐记忆的概念、测量及其实验范式。

记忆研究自古以来都备受关注,因为它与人的经验积累、才智发展密切相关。但是,对记忆进行科学的实验研究只是在 19 世纪末开始的,德国心理学家赫尔曼·艾宾浩斯(Hermann Ebbinghaus, 1850—1909)开创了这一领域,并于 1885 年发表《论记忆》一书,首次系统总结了有关记忆的实验研究成果,提出和运用了实验研究记忆的方法。艾宾浩斯的记忆研究范式一直持续到 1965 年前后才发生根本改变。1965 年,记忆研究领域出现了两个完全崭新的方向,即两种记忆(dual memory)研究和元记忆(metamemory)研究。也由此开始,记忆研究向着多种完全不同的方向前进,这些研究都深刻地丰富和发展了实验心理学。100 多年来,记忆研究领域成果丰硕,成为心理学科知识增长最快的分支之一,特别是 20 世纪中期以后,信息加工心理学开始对记忆过程的精细分析,从而使我们对记忆系统的知识存储有了更深入的理解,并结合现代认知神经科学研究

手段，推动着该领域研究的快速发展。

第一节　记忆实验研究简史

一、艾宾浩斯的开创性实验

哲学中关于记忆的研究是古老的，但采用实验方法系统地探索记忆过程及其与其他因素的函数关系，还是从艾宾浩斯开始的。艾宾浩斯受到费希纳《心理物理学纲要》的启发，决定要用自然科学的方法研究比费希纳研究的感知过程更复杂一些的心理现象。他在《论记忆》一书的序言中写道："……由下述的研究，我们试图对心理活动深入一步，把记忆的表现作实验的与数量化的处理"，艾宾浩斯这样提出研究的问题，"如果把某种无意义音节系列学习到能够背诵，搁置下来不管它，只在时间的影响或充满时间间距内的日常生活的影响下，遗忘的进程会是怎样的呢？"①艾宾浩斯"将记忆作为一个心理过程，并将其分成学习（识记）（learning）、保持（retention）、联想（association）和复现（reproduction）四个阶段来研究，他认为这样分段便于控制实验条件"。②他以自己为被试，以无意义音节（nonsense syllable）为材料，采用节省法（saving method）系统探究了时间因素在遗忘进程中的影响，绘制出经典的"遗忘曲线"（forgetting curve）。艾宾浩斯在一项研究中，设计了如下的实验程序：

他使用德文字母编制"无意义"的实验材料：在两个辅音字母之间插入一个元音字母构成无意义音节，如 RAF、NUZ、ZUP 等。以 13 个无意义音节组成一个音节组，8 个音节组作为一次实验学习的材料。在一次实验中，他规定自己要学习的音节组和每个音节呈现的速度，一遍一遍地学习直至能连续背诵两遍不出错，然后将这组学习材料搁置不管。间隔一定时间，再回忆这组音节就会出现困难，于是重新一遍一遍地学习，直至重新能连续背诵两遍不出错，比较前后两次学习所使用的时间或遍数，就可以计算出第二次学习中节省的时间或遍数。而这种节省是由于第一次学习后对材料的保持，所以艾宾浩斯使用第二次学习

① Ebbinghaus, H. (1885). *Memory: a contribution to experimental psychology*. New York: Dover.

② 杨治良编著：《记忆心理学》，华东师范大学出版社 1999 年版，第 9 页。

节省的时间或遍数的百分比来测量对第一次学习材料的保持量,故称节省法。假如他最初学会一组无意义音节耗费了 1 156 秒,间隔 20 分钟后重学时耗费了 467 秒,第二次学习节省了 1 156－467＝689 秒的时间,占最初耗费时间的 64.3%,这就意味着他第一次学习的材料在经历 20 分钟的时间间隔后依然保持了 64.3%的量。他在这项研究中使用的时间间隔有 7 种:1/3 小时、1 小时、9 小时、1 天、2 天、6 天、31 天。每次实验均使用新的无意义音节组。在这项研究中,艾宾浩斯先后学习了 1 200 个以上的音节组,共计背诵了 15 000 多个无意义音节,如此辛苦地工作帮助他绘成了经典的遗忘曲线,也叫做艾宾浩斯曲线(Ebbinghaus curve)或艾宾浩斯保持曲线(Ebbinghaus curve of retention),如图 8-1 所示。

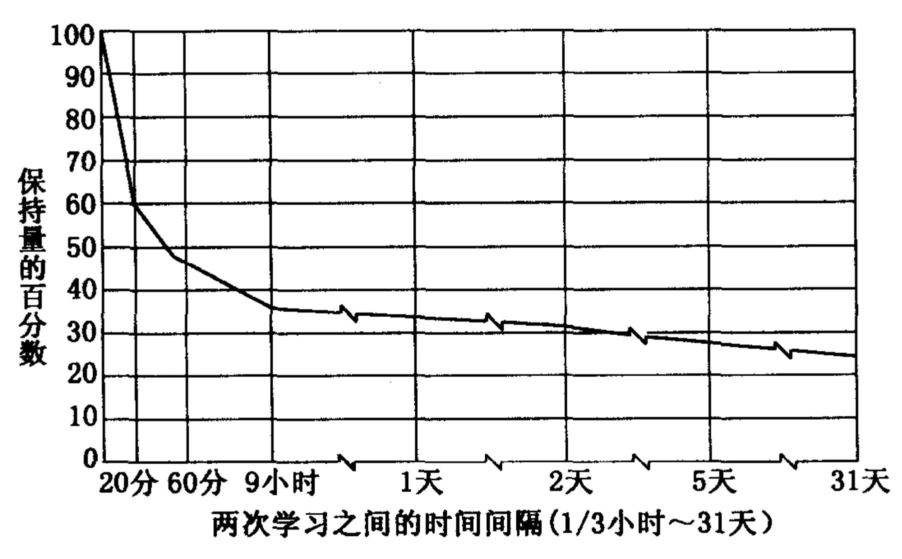


图 8-1　艾宾浩斯遗忘曲线(Ebbinghaus, 1885)

图 8-1 清楚地表明,艾宾浩斯在学习一组无意义音节后,较短时间内发生了快速的遗忘现象,而随着时间的推移,遗忘(forgetting)的速度会越来越慢,一个月后保持量稳定在 25%左右,但依然会继续缓慢下降。

艾宾浩斯的一系列实验结果得到了后来许多研究者实验结果的证实,他采用的实验方法也被其他研究者应用到其他类型材料的记忆研究中。我们说,艾宾浩斯开创了记忆实验研究的时代!尽管他使用的实验材料缺乏生态学效度,尽管他对于遗忘发生中的影响因素还缺乏全面考察,尽管他对遗忘发生的机制也还缺乏研究,但是他的名字已经并将永远地成为几乎所有学习心理学的人都熟知的名字。

艾宾浩斯之后,铁钦纳(Titchener, E. B.)也专门研究了记忆,他开展的研究是着重于对记忆表象、后象、意象类型和类似过程进行内省描述。虽然铁钦纳在

讨论中也触及当代人感兴趣的记忆研究专题，如倒摄抑制(retroactive interference)、联想干扰和中介联想等，但总的看来，他的研究几乎同当前的记忆研究不相干。①

二、巴特利特的记忆研究

自然，艾宾浩斯的早期研究也为后来者提供了质疑和批评的靶子，特别是他使用的实验材料和采取的实验手段。被誉为“英国第一位实验心理学家”的巴特利特(Frederich Charles Bartlett, 1886—1969)认为，人们的记忆是一个意义记忆的过程，而不是一个无意义音节的记忆过程，无意义音节的研究缺乏现实意义。他认为，在记忆过程中，人们会有意无意地将记忆的内容纳入到自己已有的知识结构中，也就是将新的记忆材料纳入到自己已有的图式(schema)中。

巴特利特1932年对记忆进行了一系列的实验研究，他集中研究的是记忆保持的材料怎样随时间而变化，识记材料如何被记忆保持，这与艾宾浩斯研究的侧重点不同。不过，巴特利特与艾宾浩斯记忆研究的最大差异还是在于研究材料。巴特利特选用有意义的故事和图画作为实验材料，这样的实验材料更符合人们实际的记忆活动，因此更有可能探寻到人类记忆的规律。巴特利特的研究更加强调社会文化对人的记忆过程的影响，著有《记忆：一个实验的与社会的心理学研究》(*Remembering*: *A Study in Experimental and Social Psychology*)(1932)。②巴特利特对记忆研究的贡献主要包括以下两个方面。

1. 以故事为材料的记忆实验

巴特利特最引人注目的一项研究，就是让大学生阅读然后回忆北美印第安民间故事“幽灵之战”(*The War of the Ghost*)。他采用重复再现法(repeated reproduction)，即让被试学习故事之后，按照自己的记忆讲述故事，而且至少讲述两次，两次间隔的时间中不插入学习。结果发现：被试回忆的内容和原故事的内容相比有很大差异，他们加入了许多自己主观臆想的细节。巴特利特因此认为：回忆过程中包含了许多对学习内容的重新构造成分，人们倾向于利用自己头脑

---

① 杨治良：《记忆心理学》，华东师范大学出版社1994年版，第2页。

② 巴特利特著，黎炜译：《记忆：一个实验的与社会的心理学研究》，浙江教育出版社1998年版。

中已有的图式来重新加工所获得的新信息。巴特利特由此开创了记忆错误或错误记忆的研究。

2. 以图画为材料的记忆实验

巴特利特的另一类记忆研究使用的是图画材料。实验方法是：首先让一个被试观看一幅图画，然后要求他凭记忆将这幅图画回忆并勾画出来，再将他勾画的图形交给第二个被试观看并也要求他根据记忆勾画所看的图形……如此在多名被试间传递和勾画图形。结果发现，最后一名被试勾画的图形已与第一个被试所看的图形大相径庭了。

3. 将图式概念引入记忆心理学

巴特利特将心理"图式"的概念引入到记忆心理学中，并明确提出"表征"(representation)概念。他假设"感觉信息和心理图式一起被结构化和储存，而心理图式本身被表征在记忆之中"。①很显然，图式概念能够有效地解释巴特利特在使用故事或图画为材料所做的记忆实验研究的结果。

巴特利特开辟了记忆研究的另一条道路，是对艾宾浩斯记忆实验研究的继承和拓展，他在记忆研究方面占有重要的地位。

此外，与艾宾浩斯同时代，还有许多与记忆有关的研究和属于记忆研究领域的重要概念被提出。比如在英美等国家，联想主义和行为主义范畴内的许多研究都与记忆密切相关，像巴甫洛夫、桑代克、斯金纳、赫尔、托尔曼等进行的动物学习的研究也都可归入记忆研究范畴。雅各布斯 1887 年给被试读出一串数字后要求其随即进行回忆，结果能回忆的数字个数最多在 7 个左右；詹姆斯在 1890 年提出初级记忆(primary memory)和次级记忆(secondary memory)的观点；米勒在 1900 年提到过短时记忆的概念，这些研究和概念的提出为 20 世纪中期兴起的记忆结构研究做了思想和概念的准备。

三、信息加工心理学的记忆研究

按照信息加工的观点，记忆是信息的输入、编码、贮存和提取的过程，它能更全面地体现信息加工系统的工作流程，所以它也是信息加工心理学研究的核心内容之一。到目前为止，有关记忆的信息加工研究集中在三方面：记忆系统的结

① 杨治良等：《记忆心理学》，华东师范大学出版社 1999 年版，第 10 页。

构、记忆内容的表征和记忆系统的容量，主要是记忆的结构和信息表征。

（一）记忆结构的研究

记忆结构的研究是从两种记忆说(dual memory theory)开始的。早在1890年美国心理学家詹姆斯(William James, 1842—1910)就提出初级记忆(primary memory)和次级记忆(secondary memory)的概念，前者就是指短时记忆，后者就是指长时记忆。1965年，沃和诺曼(Waugh & Norman)正式提出两种记忆说，即在人的长时记忆系统(long-term memory system)之外还存在短时记忆系统(short-term memory system)。不过，按照斯珀林(Sperling, 1960)的研究，短时记忆之外还存在保持信息时间更为短暂的瞬时记忆系统，也叫做感觉记忆系统(sensory memory system)。斯珀林使用巧妙的部分报告法实验，证实了视觉系统的图像记忆(iconic memory)，随后，其他研究者也证实了听觉系统中的声像记忆(echoic memory)。图像记忆和声像记忆对信息保持的时间非常短暂，并且信息保持本原特征，是未对信息进行加工的系统。在这些研究基础上，阿特金森和希夫林(Atkinson & Shiffrin, 1968)提出了记忆的三级加工模型。该模型认为，记忆结构是固定的，而控制过程是可变的，记忆由感觉记忆、短时记忆和长时记忆三个存贮系统组成。在信息加工过程中，外部信息首先进入感觉记忆，感觉记忆对信息保持的时间非常短，只有1秒钟左右，然后那些受到注意的信息获得识别进入短时记忆，未受注意的信息则迅速消失。短时记忆是一个信息加工的缓冲器，其中的信息处在意识活动的中心，但是其容量有限，只能保留7±2个信息组块，而且信息保留的时间也只有1分钟左右，除非不断对信息进行复述。复述可以使短时记忆中的一部分信息进入长时记忆。长时记忆的容量很大，对信息保留的时间也可以很长，它是我们的信息库，我们积累的大量知识经验都贮存在这里。长时记忆中的信息可以在激活信号的作用下回到意识状态，供认知系统应用。

对于记忆的多存贮理论也有不同意见，其中最具代表性的是克雷克和洛克哈特(Craik & Lockhart, 1972)的加工水平说。加工水平说认为，记忆系统并不存在多个存贮结构，信息保持时间的长短不是由于信息所处系统的不同，而只是由于受到了不同水平的加工。

（二）记忆信息表征的研究

记忆信息表征主要是长时记忆的信息表征，长时记忆信息也称为知识(knowledge)，即个人知识。随着近年计算机科学和人工智能的发展，知识工程

的诞生，科学家们越来越关心“知识的组织”方式。知识分为两大类：陈述性知识(declarative knowledge)和程序性知识(procedural knowledge)。陈述性知识是关于事实的信息，例如，月亮是绕着地球运行的，它是一种相对静态的知识，随时间变化较小；它是可以被描述的，即它可以被分解为一系列有关的事实。当然，在人们头脑中保存的知识也可能包括某些错误的描述。程序性知识通常是指有关技能和解决问题方面的知识，如体操运动的知识、骑自行车的知识等，它们是动态的，随时间变化较快。而且，这种知识往往不能像陈述性知识那样被很清楚地讲给别人听。另外，程序性知识在人脑中是如何组织和保存的呢？陈述性知识和程序性知识有什么关系？这两种知识在头脑中的编码方式有什么区别？这些问题仍然未被很好地解决，是属于记忆领域深层次研究的问题。陈述性知识的表征方式既有情景性的，也有语义性的(Tulving, 1972)，其中语义记忆信息的表征模型主要包括层次网络模型(hierarchical network model)和特征比较模型(feature comparison model)两类。前者认为人脑对语义的记忆是以网络形式分层存贮的，所有的概念均按照逻辑的上下级关系或类属关系分若干层次，各层次的概念依次有连线相通，由此构成一个层次网络，概念的特征附着于网络的各个结点上。后者则认为概念的表征依赖特征集，任何概念都包括一个定义特征集和一个描述特征集，两个概念的特征交叉得越多，概念的重叠就越多，关系就越密切。

(三) 内隐记忆的研究

近年来，许多心理学家对内隐记忆(implicit memory)表现出浓厚兴趣。其实，早在17世纪就有学者对内隐记忆现象进行了观察和描述。如笛卡儿(Descartes, 1649)观察到儿童期受惊吓的体验可以影响其后的一生，但他们对具体事件并无记忆；法国哲学家比朗(Biran, 1824)认为，在充分重复后，习惯可以自动地或无意识地进行，而并不觉察到动作本身或先前习惯形成阶段的情节；艾宾浩斯(1885)、桑代克(1934)等都对内隐记忆现象有观察和描述，只是没有明确使用这一术语而已。

最早使用“内隐的”(implicit)和“外显的”(explicit)这对术语来描述不同记忆形式的心理学家是麦独孤(William McDougall, 1871—1938)，他把对过去事件有意识的回忆称为外显回忆，而把没有进行有意识地努力或清晰意识到、但其行为受近期事件影响而产生某种变化的现象称为内隐的再认。只不过，他的观点当时并未引起其他研究者足够重视。直到1968年，沃灵顿和韦斯克兰茨(War-

rington & Weiskrantz, 1968, 1974)在对遗忘症患者(amnesic patient)启动效应的研究中发现:不能有意识地保持学习内容的遗忘症患者,尽管在再认测验中不能辨别出先前学习阶段呈现过的单词,但在补笔测验中却对先前呈现过的单词表现出正常的保持效果。这一结果激发了人们对无意识的、无察觉的记忆现象的研究兴趣,也激发了人们对有意识记忆测验和无意识记忆测验进行系统比较研究的兴趣。在更进一步的研究中,采用的测量方法一般不要求被试回忆或再认,即不要求有意识地提取过去呈现的特定信息,而是要求被试完成一系列的相关任务,通过被试对这些任务的操作来反映其对先前学习阶段获得的信息的记忆效果。这类不需有意识"回忆"的记忆任务称作内隐记忆(implicit memory),而对先前信息的有意识回忆则被称为外显记忆(explicit memory)。内隐记忆和外显记忆的区分和研究是记忆领域中的一个新课题,其中还有许多疑点尚未查明。

四、认知神经科学中的记忆研究

认知神经科学(cognitive neuroscience)是当前实验心理学的前沿领域,它试图探索伴随各种心理过程而出现的神经运动过程,甚至神经原因,这其中也包括记忆的神经机制研究。关于记忆的认知神经科学研究主要包括以下几个方面。

(一) 记忆是否存在脑区定位

许多研究显示,记忆与脑的特定部位有关,不同类型的记忆在脑的不同部位产生。大脑两半球的不对称性在记忆方面也有突出表现。左半球言语运动中枢受损伤,不仅导致失语症(aphasia),而且损伤言语记忆,病人可以记住事物的形象及属性,却不能记住其称名词汇。相反,右额叶损伤后,直观形象记忆发生困难,然而言语记忆几乎不受影响。加拿大著名神经外科医生潘菲尔德(Wilber Graves Penfield, 1891—1976)和他的同事在医治癫痫病人时的大量材料也给定位说以有力的支持。他们认为记忆与大脑额叶和颞叶有密切关系,因为他们发现当额叶损伤严重时,病人不仅会丧失有组织的逻辑思维能力、计划能力,而且也不能形成稳定的动机和有目的的回忆。科恩(Cohen)在给抑郁症患者脑区不同部位以电击产生痉挛时,发现电击左脑会对言语记忆造成损害,但对形象记忆几乎没有影响;相反,电击右脑则对形象记忆造成损害,却对言语记忆影响不大。由此推断,言语记忆可能储存在大脑左半球,形象记忆可能储存在大脑右半球。

潘菲尔德曾多次报告过,他在给重度癫痫病人施开颅手术时,用微电极刺激患者右侧颞叶,病人忽然产生对"遗忘"往事的鲜明回忆,甚至听到了过去曾听到过的歌曲,能随着音乐节律把歌唱出来。他称这种记忆为"诱发回忆"(evoked recall),诱发回忆大多数是视觉和听觉方面的。而电刺激大脑皮层其他部位时,就没有这类反应。

前苏联神经心理学家鲁利亚(Александр Романович Лурия, 1902—1977)发现皮层下组织与记忆有关。海马、乳头体和边缘系统受损伤病人的短时记忆出现明显障碍。米尔纳(Brenda Milner, 1918—)的实验还证实了海马在从短时记忆向长时记忆的信息转化中起着非常重要的作用。上述事实表明,海马、乳头体、丘脑背内侧核可能是形成、存储短时记忆信息的特定部位,而两侧额叶可能是存储长时记忆信息的重要部位。

另有学者认为,记忆是一种复合心理现象,大脑中可能不存在单纯的记忆中枢。美国神经心理学家拉什利(Karl Spencer Lashley, 1890—1958)最早提出记忆与大脑的各个部分都有关而无特殊定位,这种主张被称为脑均势说。他在动物身上进行一系列实验,破坏动物大脑皮层的不同区域,并检查其与记忆学习的关系,发现大脑皮层破坏的区域越大,学习受到的影响就越严重,记忆丧失也越严重。因此,他认为记忆的保持,不依赖大脑皮层的精细定位,而是整个大脑皮层的机能。这一主张得到了20世纪60年代兴起的"聚集场"假说的支持,这种假说认为,神经细胞都不能离开细胞通路系统,任何一个神经细胞都不能离开细胞群而单独活动。一个神经细胞,可以是这一通路上的一个环节,同时也可以是另一通路的组成部分。因此,记忆痕迹并不是依靠某一固定的神经通路,它涉及成千上万甚至上百万神经元的相互联系。

从大量已有的实验结果及临床经验看,定位说和均势说均有证据,还都无法被否定。应该看到,记忆是整个神经中枢系统的功能,是中枢神经系统不同部位参与的联合活动,但不同部位所起的作用存在差异。信息的储存多数在大脑皮层,但皮层下部位,如丘脑、边缘系统、脑干网状结构甚至脊髓,也都具有存储信息的功能。

(二)突触变化是记忆的生理基础吗

澳大利亚神经生理学家埃克尔斯(John C. Eccles, 1903—1997)使用独特的实验技术,对突触及单个脑细胞的电生理活动进行了大量研究,他发现刺激的持

续作用可使神经元的突触发生变化，如神经元的轴突末梢增大、树突增多变长、突触间隙变窄、突触内的生化变化使相邻神经元更易于相互影响等。突触的这些变化使信息传入效率大大提高。这一看法得到不少实验的支持。例如，将刚生下的一窝小白鼠分为两组，一组饲养在有各种设备和玩具、内容丰富的环境里，另一组放在没有任何设备的贫乏的环境里，30天后，发现前一组白鼠的大脑皮层在重量和厚度上均超过后一组白鼠，它们的突触数目也增多，脑中与学习有关的化学物质的浓度较高，学习行为表现较好。因此，有人认为突触结构的变化，可能是长时记忆的生理基础。

（三）生物大分子是信息的储存库吗

20世纪60年代，分子生物学兴起，生物大分子在大脑活动过程中的作用的研究有较大进展，这就为在分子水平上揭开记忆之谜打下了基础。已经发现，脱氧核糖核酸(DNA)借助核糖核酸(RNA)传递遗传信息，一些心理学家因此假定，个体记忆经验是由神经元内部核糖核酸的分子结构来承担的。由学习引起的神经活动，可以改变与之有关的那些神经元内部核糖核酸的细微化学结构。获得诺贝尔奖的瑞典神经生物化学家海登(Hyden, H.)首先进行了这方面的研究，他训练小白鼠走钢丝，然后对其进行解剖，发现鼠脑内与平衡有关的神经元的RNA含量显著增加，而且组成成分也有变化。据此，海登认为大分子是信息的储存库，RNA和DNA均是记忆的化学分子载体。

后来，有研究者将抑制RNA产生的化学物质注射到动物脑内，结果会使该动物学习能力显著减退或完全消失；而将能促进RNA生成的化学物质注入动物脑内，则能提高动物的学习能力。这进一步证明RNA本身的变化是学习和记忆的重要物质基础。后来海登与美国的弗莱克斯夫妇先后发现，记忆与RNA合成蛋白质有关。学会走钢丝的小白鼠，由于注射了嘌呤霉素，干扰了它的RNA合成蛋白质，于是小白鼠忘记怎样走钢丝。此时RNA没有减少，只是合成蛋白质出现了困难，因而造成记忆进程缓慢。RNA不但影响蛋白质的合成，也影响神经元之间递质的释放，因此影响记忆过程的进行。

认知神经科学对于记忆的实验研究才刚刚起步，还有不少观点尚属假说，仍需检验，但这些初步的研究成果，使我们对记忆本质的理解更加深入。可以预期，随着现代科学技术的进步及科学研究的集成化，人类揭开记忆的神经之谜已为期不远了。

第二节　记忆实验研究的基本方法

传统实验心理学把记忆划分为三个阶段，即识记、保持、回忆或再认。自然，记忆的实验过程一般也被划分为三个阶段：材料的呈现与学习、一定的时间间隔、保持量的检查。研究者通常都是将研究变量结合在这三个阶段的过程中，考虑的问题主要有：被试因素，如年龄、性别、健康状况、教育程度、认知方式、信念和价值观、动机系统等；学习材料的性质、难度、总量及呈现方式，如言语或图形材料、简单或难学的材料等；时间间隔的设置及插入其中的其他因素，如较短的或较长的时间间隔、即时回忆(immediate recall)或延迟回忆(delayed recall)、时间间隔中插入干扰因素或不加入任何干扰；保持量(retention value)的检查方式，如回忆法(recall method)、序列重建法(series reconstruction method)、节省法(saving method)或再认法(recognition method)等。

一、实验材料的选择与编制

可用的记忆实验材料十分丰富，从理论上说，凡是能被人们学习的事物都可以成为记忆实验的材料，如语言文字的、图形的、场景的等等。不过，为便于实验操作，实验心理学中一般要求记忆材料的选择与编制要符合以下标准。①

第一，对于每个被试来说，实验材料在熟悉度上应基本相等，最好对于所有被试来说都是新颖、不熟悉的。这样才能保证所有被试都是从同一基线开始学习，最后观测到的保持效果才能归之于实验过程中学习的效果和实验条件的控制作用。

第二，使用的材料，在数量上或意义上容易被分割成相等的、较小的单元，以便将被试在保持量检查阶段提取的资料数量化。当然，这不是绝对的，若以掌握全部材料需要的时间或尝试的次数作为记分标准时，就无需把材料分割成相等的单元。

第三，使用的材料，应是被试能在较短时间内学会和掌握的。如果一次使用的材料太多、费时太长，势必会影响实验的效果，同时把被试长时间留在实验室

① 杨治良：《实验心理学》，浙江教育出版社1998年版，第468页。

内也是不可能的。

根据以上标准，在记忆实验中可选用的材料仍是很多的。这些材料概括起来有两大类：言语材料和非言语材料。

言语材料是指语音音节、字母、单字、单词、句子、诗歌、散文和数字等。这些材料又可分为无意义的和有意义的两种。如无规则组成的音节、无规则排列的字母和随机数组都属于无意义的言语材料，而短语、句子、诗歌和散文等则属于有意义的言语材料。

运用言语材料对记忆进行系统的实验研究，始于艾宾浩斯。他为了排除联想的影响，首创了无意义音节。但是，无意义音节只是对被试熟悉的语言来说没有意义，并不意味着被试对这些音节就真的不产生任何联想，有些音节的联想值还可能很大。由于人们对联想值大的音节的学习速度比联想值小的音节的学习速度快，因此在使用无意义音节作为实验材料时，最好对使用的音节加以定标，即预测一下每个音节联想值的大小，以确定其难易和取舍。测定联想值的方法是分别把每个音节呈现给若干被试，令其在一定时间内（一般为 2～3 秒钟）说出对这个音节产生的联想，然后根据对每个音节产生联想的人数多少来计算各音节的联想值。联想值可用百分数表示。

非言语材料主要指各种实物、图形和“感知-动作”性的学习工具（如迷宫、镜画、积木、打字、投球等）。这类材料在学习和记忆实验中都是常用的，尤其是在研究学前儿童和动物学习、记忆实验中使用得更为频繁。

在人类记忆的实验中，上述两类材料的区分只是相对的，因为人有第二信号系统。人的第二信号系统是与第一信号系统协同活动的，即使在学习和记忆一些非言语材料时，也离不开言语的活动。例如，在迷宫的学习活动中，被试的整个操作过程虽然主要依靠触觉和动觉，但是也伴随有言语活动，常常是用言语来支配和调节自己的动作。因此在使用非言语材料进行实验时，也应考虑到言语活动的因素。

二、实验材料的呈现

在实验中，把选定的材料用什么方法向被试呈现，是需要慎重考虑的问题。因为呈现材料的方法本身就是一个可研究的变量，不同的呈现方法会导致不同的学习方式，也会产生不同的实验效果。记忆心理学中常用的学习材料呈现方

法包括全部呈现法、提示法、对偶联合法等。

(一) 全部呈现法

这种方法是把需要被试学习和记忆的材料一次性全部呈现给被试，让其识记。在具体使用这种方法时又可分为两种情况：一种是限定时间的学习，要求被试在规定时间内学习全部材料，然后测定其掌握的程度；另一种是不限定时间的学习，只是要求被试学会到某一标准时停止，根据被试学到规定标准需要的时间长短来衡量其记忆效果。

全部呈现法的优点是：简便易行，不需要什么特殊的仪器设备就可进行；能充分发挥被试学习的主动性，使被试可根据材料的数量和性质自由地运用一些学习方法进行学习和记忆。因此，在研究不同被试对同一材料或同一被试对不同材料的学习方法时，这是一种较好的材料呈现方式。当然，通过这种呈现方式获得的数据，只能表明被试在一定时间内学会多少项目或学会全部材料需要多少时间，而显示不出被试对材料的各个项目和全部材料的学习进程。

(二) 提示法

提示法(anticipation method)又称为预测法或系列学习法。这种方法是把材料的各个项目按一定的速度依次向被试呈现，让被试学习和记忆。第一遍呈现时，要求被试按顺序记住各个学习项目；从第二遍开始则要求被试在每个记忆项目呈现时尽量正确地说出下一个即将出现的项目是什么，不管其报告的是对还是错，下一个项目都会按照事先规定好的速度呈现出来，这种呈现方法可以对被试的报告起到一种强化、纠正和提示的作用。如此，一遍一遍地进行下去，直到被试能无误地全部报告出整个刺激系列为止。

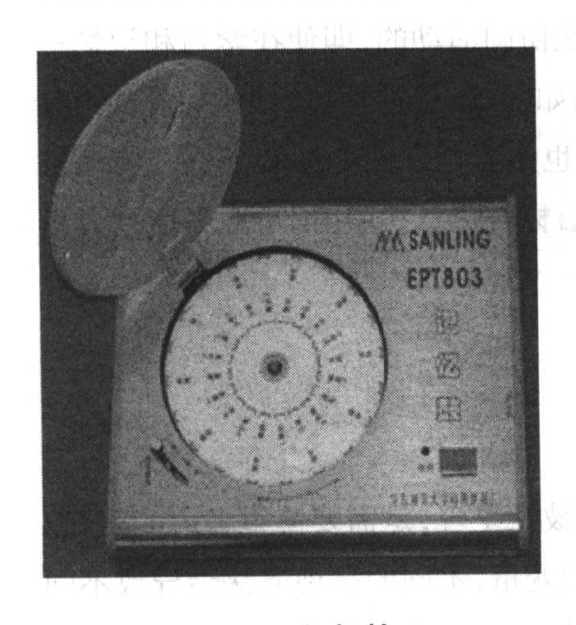


图 8-2　记忆鼓

在记忆实验中，为了让实验者能更好地记录被试的反应，通常要借助一些仪器设备呈现学习材料和记录被试的表现。记忆实验室常备的实验装置主要包括投影机、摄像机、录音机、计算机、速示仪(tachistoscope)、记忆鼓(memory drum)等。其中的记忆鼓算是专门的记忆研究装备，这种装置一般包括三个主要部分：带孔(或称窗)的挡板、挡板后面的圆纸盘、由电机自动控制的金属鼓，如图 8-2 所示。

实验中，把要被试学习与记忆的系列材料印制在圆形纸盘上，安装在电动控制的金属鼓上，盖上挡板，设定金属鼓的转动速度和在每个项目上停留的时间。金属鼓带动纸盘使其逐段逐段地移动，被试就可以从刺激窗孔看到一个又一个的学习项目，照此循环下去就可以获得被试学习进程的研究资料。当然，记忆鼓的具体式样是多种多样的。

图 8-3 是说明采用提示法学习一个无意义音节序列的实验程序。①图的中间一列是要被试学习的 12 个无意义音节，左边一列数字是对应的每个音节在刺激序列中所处的位置，也是呈现的顺序号，右边一列音节则是在刺激项目呈现时，被试应报告的下一个将会呈现的音节。实验开始后，先从起始的提示信号(XXX)开始，以 3 秒钟呈现一个项目的速度向被试呈现全部学习系列，然后再从提示信号开始进行第二遍学习，呈现速度也是每个项目 3 秒钟，从这一遍开始要求被试在每个项目呈现时说出与其相关联的下一个项目。例如，当提示信号呈现时，被试要说出第一个音节“HIG”；而当第一个音节(HIG)呈现时，被试要说出第二个音节“WUG”，如此一遍一遍地进行测试，直到被试能全部正确地报告出每个音节为止。

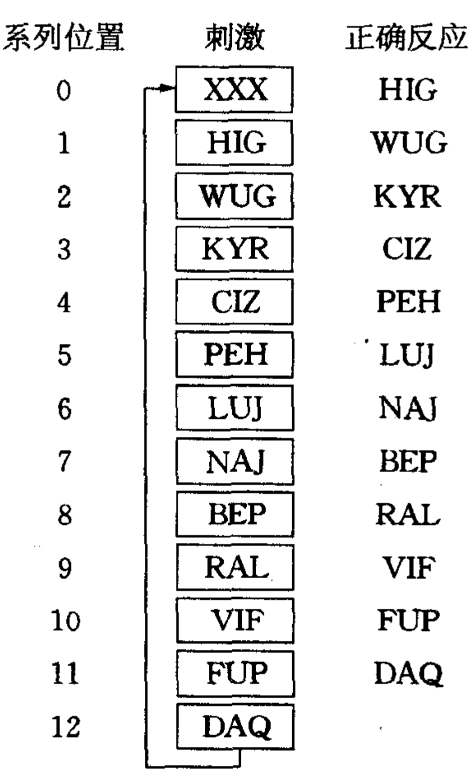


图 8-3 系列学习的提示程序(杨治良，1988)

采用提示法，可以有效地控制每个项目呈现的时间，以及整个刺激系列呈现需要的总时间，而且因为每学习一遍的同时也对被试达到的学习程度进行了一次测试，所以可以利用这种实验的结果绘制出学习的进程曲线，即学习曲线(learning curve)，借此了解系列位置对记忆的影响。

(三) 对偶联合法

对偶联合法(paired association method)是在记忆实验中将刺激材料成对组合，要求被试学习时记住一对刺激项目的对应关系，如图 8-4 所示。成对材料中

① 杨治良：《基础实验心理学》，甘肃人民出版社 1988 年版，第 288—289 页。

学习的对偶项目	检查学习效果
铁路——教师	橘子——
黄山——电视	数目——
北京——教授	鲜花——
橘子——桃子	铁路——
蔬菜——大米	蔬菜——
广州——课本	黄山——
鲜花——太阳	火车——
围裙——战争	北京——
数目——动作	围裙——
火车——毛皮	广州——

图 8-4　对偶联合法实验材料示意图

的前一个项目又可以叫做刺激项目，后一个项目又叫做反应项目，因为在记忆效果检查过程中一般是给出前一个项目，要求被试报告出其对应的另一个项目。在采用这种方法的实验中，主要是要求被试记住两个项目间的对应关系。就像记忆英语单词一样，它要求学习者将一个中文词汇与一个英文字母串联系起来，建立二者的对应关系。图 8-4 中的两个词汇组成一个对偶，其中前一个是刺激词，后一个是反应词。学习时向被试呈现成对的两个词，然后呈现每对词的刺激词，要求被试报告对应的反应词，比如主试给出“黄山——”，则被试就应给出“电视”；主试给出“广州——”，则被试就应给出“课本”，否则就是错误的。

采用对偶联合法进行记忆实验时，又可以使用两种方法检查被试的保持量，一种是“回忆法”，另一种是“提示法”。图 8-4 所示的就是回忆法，即当成对的材料系列学完后，将所有材料中的“刺激项目”呈现给被试，被试回忆每一个刺激词对应的反应词；如果采用提示法，则是在第一遍学习后，从第二遍开始，每当呈现出一个刺激词时，都要求被试回答对应的反应词，而且不管其是否正确地回答出反应词，这个反应词都会按照既定的时间再呈现出来，如此一遍一遍地进行下去，直到完全学会为止。如前所述，采用这种提示法有利于控制被试学习的时间，有利于测量被试的学习进程，也可以进行项目难度分析等。

三、保持量的检测

在记忆实验中，检查记忆效果的方法很多，而且保持量的检查方法与刺激材料呈现的方法是相互关联的。记忆实验中常用的保持量检查方法有回忆法、序列重建法、节省法、再认法、反应时间法等。

（一）回忆法

回忆法(recall method)又称再现法或重现法，在被试识记了某一系列材料后，间隔一定的时间，然后让其把识记的材料以口头或书面的形式重现出来，以

回忆的正确百分数作为被试对学习材料的保持量的测试结果。

回忆法又分为两种:无凭借回忆和有凭借回忆。前者是对被试要回忆的材料不加任何提示,只要求被试把识记过的材料写出或说出来;后者则向被试呈现部分识记过的信息,要求被试以此为凭借回忆出对应的其余材料。①以上所讲的"提示法"、"对偶联合法"均可算作是有凭借回忆的方法。

回忆法测试得到的保持量等于被试正确回忆的项目数占全部学习项目数的百分比,即:

$$保持量=\frac{正确回忆项目数}{全部识记项目数}\times 100\%$$

(二) 序列重建法

序列重建法就是先向被试呈现一个刺激序列,给予一定的学习时间,而且要求被试记住刺激项目间的位置关系,然后将这些项目的顺序打乱再呈现给被试,让被试根据自己的信息保持恢复刺激项目的原有顺序。这种方法一般适用于非文字的实验材料,因为非文字的实验材料不便于使用口头报告或书面报告。图8-5所示的过程能够说明序列重建法实验的程序:将刺激项目按照确定的顺序呈现给被试,然后将顺序打乱以形成"任务序列",被试按照保持的信息尽量将"任务序列"中的项目顺序恢复成如"刺激序列"的顺序,被试恢复的结果叫做"重建序列",重建序列中可能会存在错误,错误越多说明记忆效果越差。

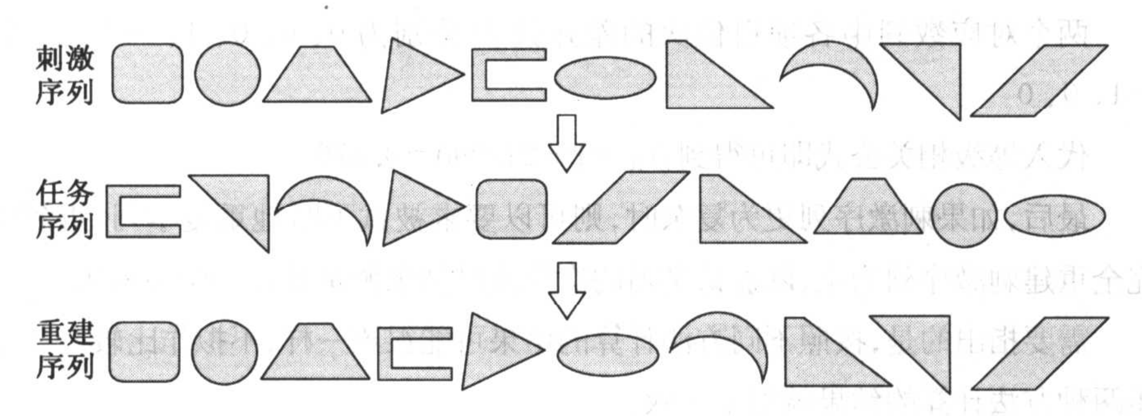


图 8-5 序列重建法实验过程示意图

序列重建保持量的计算可以有三种方法:重建正确顺序百分数法、等级相关法和学习次数或时间法。从图 8-5 中看到,被试重建序列与原来的刺激序列相

① 杨治良:《基础实验心理学》,甘肃人民出版社 1988 年版,第 293 页。

比存在差异，如果采用最简单的重建正确顺序百分数方法，则利用下列公式计算其保持量：

$$保持量=\frac{重建序列中正确的顺序数}{原刺激序列中的全部顺序数}\times 100\%$$

图 8-5 中，被试重建正确的顺序数只有 3 个，而刺激序列的全部顺序数是 9 个，所以该被试学习的保持量等于 33.3%。

其次，可以采用等级相关法，即计算各个刺激项目在原序列中的位次数列与在重建后其位次数列的相关系数，该相关系数的大小反映了重建序列与原刺激序列的近似程度，当其达到 1 时，就说明重建序列与刺激序列完全一致。等级相关[也称为斯皮尔曼相关(Spearman correlation)]的计算公式如下：

$$r_R = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

式中 D=对偶等级之差，即同一个刺激项目在两个序列中的位次之差，如图 8-5 中倾斜的三角形在刺激序列中位次是 4，在被试重建序列中的位次是 5，则二者的位次之差 $D=1$ 或 -1；n 是刺激序列中的项目数，在图 8-5 所示的实验中 $n=10$。

图 8-5 中刺激序列中各项目位次数列是 1、2、3、4、5、6、7、8、9、10

重建序列中与原序列各项目位次数列对应的是 1、2、3、5、4、6、8、7、9、10

两个对应数列中各项目位次的差异数 D 分别为 0、0、0、1、−1、0、1、−1、0、0

代入等级相关公式即可得到：$r_R=1-24/990=0.876$

最后，如果刺激序列更为复杂时，则可以要求被试不断地重复学习，直到能完全重建刺激序列为止，以被试学习的次数或时间来衡量被试的学习效果。

需要指出的是，按照不同方法计算的结果可能很不一样，不拟作比较。如上述两种方法计算的结果就很不一致。

(三) 节省法

节省法亦称重学法，就是艾宾浩斯采用的保持量检查方法。其计算公式为：

$$保持量=\frac{初学时的时间或次数-重学时的时间或次数}{初学时的时间或次数}\times 100\%$$

节省法对保持量的检查相对来说更为灵敏，因为人们对学习过的材料，经过

相当长时间后，即使不能正确地进行回忆和再认，但也仍有一定程度的保持量，这一部分的保持量成为重学时的基础，它使得重学时需要的时间或次数减少。这就是说，节省法测量到了那些未达到回忆阈限的信息项目。

（四）再认法

再认法的操作过程是：让被试学习一系列项目，然后将这些项目与一些未学习过的项目混合在一起，再一一呈现给被试，要求其对项目进行再认。这种方法在具体运用上，又可以有许多不同的变式，例如，相互混合的新旧材料的数量、相似程度均是可以改变的，而且在材料的呈现方面，可以把混合后的材料一个一个地依次呈现，要求被试判断是否学过。一般来说，混入的新材料的数量越多、其与原学习项目的相似程度越大、呈现的时间越短，再认的难度也就越大。

再认法计算保持量的公式有两种：

第一，在原识记的项目与混入的新项目数量相等的情况下：

$$\text{保持量}=\frac{\text{正确再认数}-\text{错误再认数}}{\text{原识记项目数}+\text{混入的新项目数}}\times 100\%$$

第二，在原识记的项目与混入的新项目数量不相等的情况下：

$$\text{保持量}=\left(\frac{\text{原识记项目正确再认数}}{\text{原识记项目数}}-\frac{\text{新混入项目错误再认数}}{\text{新混入项目数}}\right)\times 100\%$$

这一公式显示的计算方法实际上就是：被试的击中概率减去虚报概率。

此外，再认法的实验结果也可以使用信号检测论方法进行处理，即根据被试再认时的击中概率、虚报概率计算被试的辨别力指数和似然比，辨别力指数反映了被试对识记项目记忆的牢固程度，似然比则反映了被试在再认测验时判别标准的高低。

采用信号检测论方法来处理再认实验的结果更精确、可靠，因为传统的再认法得到的分数是被试再认辨别力和猜测成分的混合物，但信号检测论方法可将再认能力与猜测成分及辨别标准分离开来，比较纯粹地反映被试对测试项目的辨别能力。

（五）反应时间法

反应时间法（reaction time method）是根据被试对测试项目的反应速度来测试其对识记项目的巩固程度，因为这种巩固程度与项目重现（包括回忆和再认）的速度有密切关联。显然，对识记项目的保持越牢固、越清晰，反应也就越快，因

此反应时间也可以作为检查记忆效果的指标之一。如果出现被试在不同实验条件下正确再现的项目数量相等时,反应时间方面则可能提供二者的差别。

最后,需要指出的是,上述五种保持量的检查方法对保持量检查的程度有所不同,所以同样的记忆效果,采用不同的测试方法就会得到完全不同的结果,这些结果之间几乎没有可比性。我国著名心理学家陆志韦教授早在 1922 年就对此问题进行了专门的测试和比较,得到了如图 8-6 所示的结果。

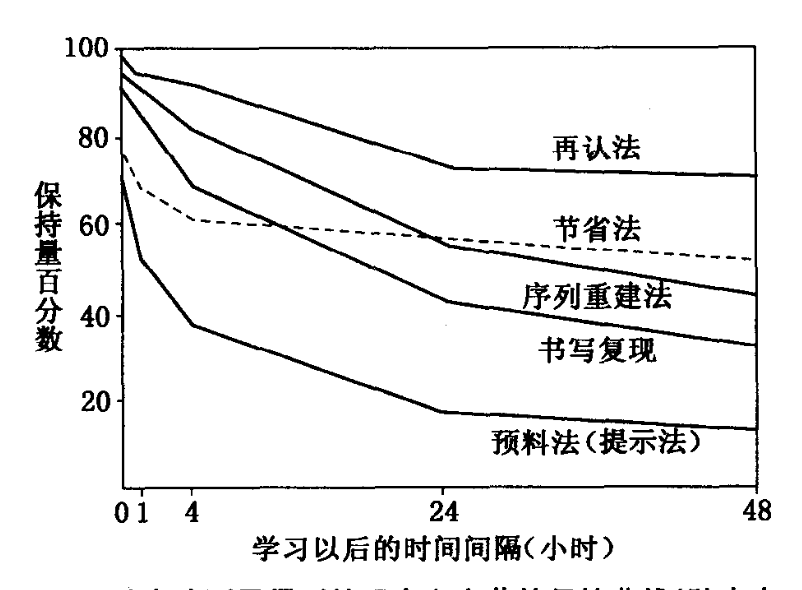


图 8-6　用五种方法测量得到的无意义音节的保持曲线(陆志韦,1922)

在陆志韦的研究中,我们看到不同保持量的检查方法所得保持量的稳定性存在差异,这意味着不同的信息提取方法对信息储存的方式提出了不同的要求,而不同方式储存的信息的牢固性也因此存在很大差异,显然,它们依赖的心理机制也可能有所不同。相对而言,再认法和节省法得到的信息保持量比较稳定,再认法对储存信息的精确性要求较低,更多的是一种整体特征的把握;节省法则可以充分地测量到那些未达到回忆阈限的信息。

第三节　记忆结构的实验研究

一、短时记忆存在吗

早在 1890 年,詹姆斯就提出记忆结构的多成分性,但这一观点在后来的半个世纪中都未受到心理学家的注意,20 世纪前半叶的记忆研究都是沿着艾宾浩

斯开拓的长时记忆研究范式展开的。到 20 世纪 50 年代后期,临床心理学和认知心理学不断提出证据证明:在长时记忆的结构之外,可能还存在另外一种完全不同的记忆系统。

1954 年,美国神经心理学家米尔纳为一名癫痫病患者实施了海马切除手术,成功地治好了他的癫痫病。但是,这位患者在术后却出现了一系列特殊的遗忘现象,比如他能毫无困难地回忆起手术前的各种丰富经历,也能像正常人那样即时回忆刚刚经历的各种生活事件,而且只要采用复述方法,他也具有很好的信息保持能力,但当他不再对信息进行复述的时候,这些信息就会很快被忘掉,即这位患者无法将新的经验保持较长的时间(Milner, Corkin & Teuber,1968)。很明显,患者的长时记忆是基本完好的,短时间内信息的保持也是正常的,他缺少的似乎是不能顺利地将短时间内的信息输送到长时记忆中去。此类现象也在其他案例中表现出来,甚至在动物实验中也都得到了验证。这些研究使心理学家重新想到了詹姆斯的初级记忆与次级记忆概念,从而开始了两种记忆理论与实验的研究,其中最经典和最直接的证据来自自由回忆实验(free-recall experiment)中关于系列位置效应(serial position effect)的研究。

自由回忆实验的基本操作是:先按一定顺序向一个或一组被试呈现一系列学习项目,然后要求被试尽量回忆出这些项目,而且不必按原先呈现的顺序来回忆。将被试自由回忆的结果与学习项目进行对照,就可得到原刺激系列中各位置刺激的回忆正确率。以刺激系列的位置为横坐标、被试对每个项目正确回忆的人数比率为纵坐标,就可得到如图 8-7 所示的自由回忆实验的系列位置曲线。默多克(Murdock, 1962)在自由回忆实验中选用的材料是 30 个单词,呈现速度是 1 秒钟呈现 1 个单词,其结果绘制成了图 8-7 所示的系列位置曲线。

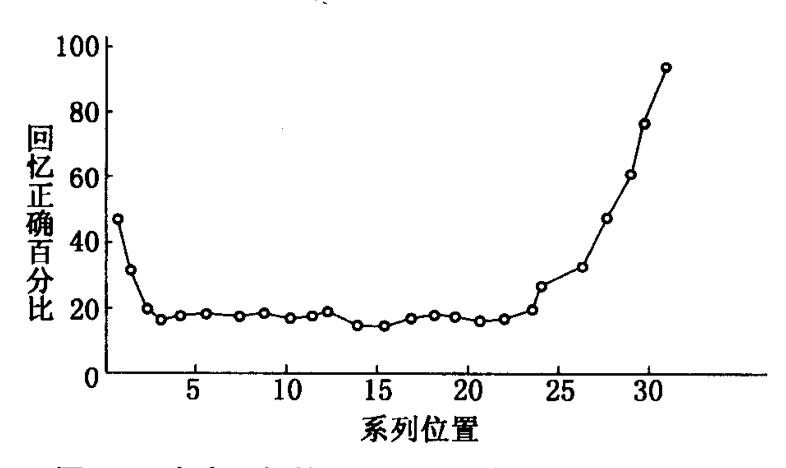


图 8-7 自由回忆的系列位置曲线(Murdock, 1962)

由图 8-7 看到，刺激系列起始部分和结尾部分单词回忆的正确率较高、中间部分单词回忆的正确率较低，这种现象就是所谓的系列位置效应。起始部分单词的回忆正确率较高，叫首因效应(primacy effect)；结尾部分单词的回忆正确率较高，叫近因效应(recency effect)。那么为什么会出现这种系列位置效应呢？根据遗忘的干扰说(interference theory)，中间部分项目的学习与回忆既受到系列起始部分刺激的干扰(前摄抑制)，又受到结尾部分刺激的干扰(倒摄抑制)，即受到的干扰总量较大，所以被试回忆的正确率较低。然而，两种记忆说却对此有着不同的解释，该理论认为自由回忆的系列位置曲线实际上可以分为两部分：起始加中间部分、结尾部分，这两部分分别反映了两种不同记忆系统的效果，其中起始加中间部分反映的是长时记忆效果，结尾部分反映的是短时记忆效果。类似研究均能证实，系列位置曲线中近因效应部分对应的刺激项目数恰好与短时记忆的容量基本一致。依照两种记忆说，曲线结尾部分的回忆成绩之所以最好，就是因为被试在开始回忆时，这一部分对应的项目尚保持在短时记忆系统中。

为了进一步证实两种记忆说的观点，心理学家就系列位置曲线提出了两点推论。

推论 1：在系列材料的呈现与学习中，如果增加刺激项目的呈现时间，也就增加了每个项目的复述(rehearsal)时间，这可使较多信息进入长时记忆，相应的回忆正确率就会提高，但项目呈现时间不会对短时记忆产生明显影响。这反映在系列位置曲线上，起始部分和中间部分会抬高，但近因效应部分不会受太多影响。

推论 2：在系列材料的呈现与学习中，如果将即时回忆改为延迟回忆并避免被试在延迟期间的复述，那就会将刺激系列最后的几个项目从短时记忆中挤出去，消除系列位置曲线上的近因效应部分，但因不改变项目呈现的时间，所以它不会对长时记忆产生明显影响。这反映在系列位置曲线上，起始部分与中间部分变化不大，但近因效应会消失。

为验证推论 1，默多克(Murdock, 1962)以刺激项目呈现的时间为变量进行了自由回忆实验。他在实验中分别以两种不同的速度向两组被试呈现单词，其中一组被试接受的项目呈现速度是 1 秒钟呈现 1 个，另一组被试接受的项目呈现速度是 2 秒钟呈现 1 个。实验结果如图 8-8 所示：两种呈现速度下都出现了系列位置效应，每个项目呈现时间为 2 秒时刺激系列的起始部分和中间部分的回忆成绩比呈现时间为 1 秒的要好，但二者在结尾部分却无明显差异。这说明增加刺激呈现时间有利于长时记忆，而不影响短时记忆。还有一些自由回忆的

实验也提供了类似证据,短时记忆的特征与长时记忆的特征也存在某些差别,如容量、保持时间和机能等。默多克等人的自由回忆实验验证了推论1。

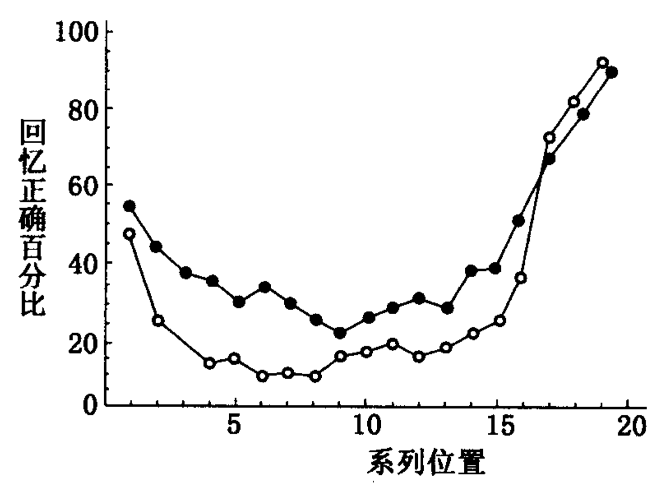


图 8-8　两种刺激呈现速度的系列位置效应(Murdock, 1962)

格兰泽和库尼茨(Glanzer & Cunitz, 1966)用实验证实了第二个推论。他们在系列单词呈现完毕后,让两组被试分别按照两种不同的方式进行回忆:一组采用即时回忆,即系列刺激呈现完毕后,立即进行回忆;另一组采用延迟回忆,即系列刺激呈现完毕后,先完成30秒钟的数字减3的逆运算心算作业,以防止复述,心算30秒钟后马上回忆前面学习的系列项目。实验结果如图8-9所示。

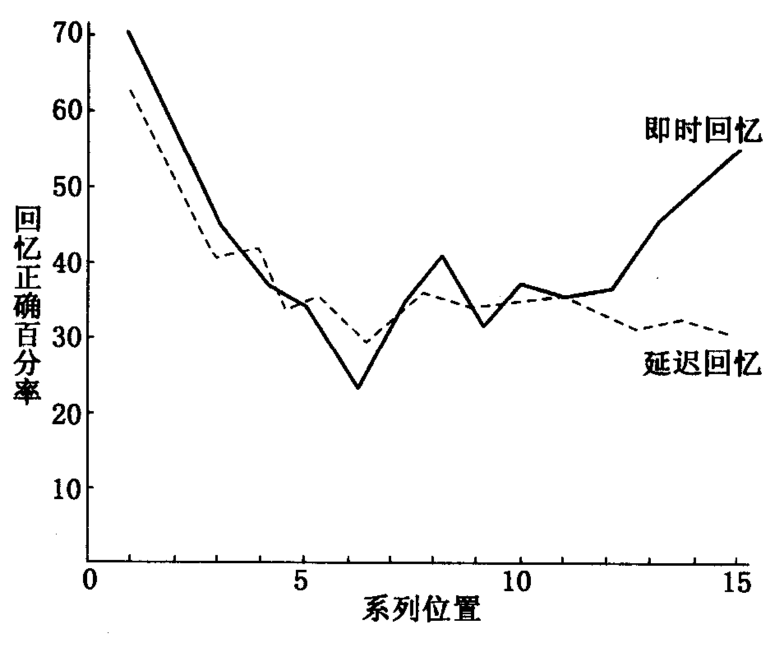


图 8-9　延迟回忆对近因效应的影响(Glanzer & Cunitz, 1966)

比较两种回忆方式下的实验结果,格兰泽等人发现即时回忆和延迟回忆都

有首因效应，而且在刺激系列的起始部分和中间部分，两种回忆的成绩均相近，这说明心算作业并未影响长时记忆。但在结尾部分，即时回忆有近因效应，而延迟回忆的近因效应消失，说明延迟回忆破坏了刺激系列结尾部分的短时记忆，这部分信息因进行心算作业而未得到复述，并被心算作业的内容所排挤，即心算作业影响了短时记忆。

两种记忆说还得到负近因效应(negative recency effect)的支持。克雷克(Craik, 1970)让被试参加一系列自由回忆实验，每次都得到一条系列位置曲线。当这些实验完成后，出乎意料地要求被试将所有实验中的单词都尽可能回忆出来，结果得到了负近因效应：在每一单个系列学习后进行的即时回忆中，系列刺激的最后几个项目都出现了近因效应，但这些项目在整个实验最后阶段的回忆测试中出现了非常低的回忆正确率。研究者解释说，这是因为在每个系列呈现完毕时进行即时回忆，最后的几个项目可以依靠短时记忆进行提取，所以被试几乎未对这些项目进行复述，其进入长时记忆的信息量自然比较少，使这些项目在最后的回忆率非常低。①

短时记忆的存在得到许多研究的证实，记忆研究因此发生了革命性变化，出现了两种记忆说。两种记忆说认为记忆不是单一的东西，它存在彼此独立而又相互联系的短时记忆和长时记忆两个系统，二者共同构成了一个统一的系统。因此，这种理论认为记忆结构是二分的，其对信息的加工过程是：外部信息经过感觉通道先进入短时存贮器，短时记忆容量以内的信息可在短时存贮器中短暂保持约15～30秒，但如对这些信息进行复述，则信息可随复述而继续保持，并可能进入长时存贮器。因此，短时存贮器是信息进入长时记忆的一个容量有限的缓冲器和加工器。

二、部分报告法实验与感觉记忆

1960年，斯珀林首先提出了瞬时记忆(immediate memory)或感觉记忆(sensory memory)的概念，并使用巧妙的部分报告法实验证实了这种记忆的存在。该研究成为三种记忆模型的经典实验证据和基础。

(一) 部分报告法实验与图像记忆

视觉的瞬时记忆现象，在心理学中早有研究。1859年，汉密尔顿(Hamilton,

① 鲁忠义、杜建政：《记忆心理学》，人民教育出版社2005年版，第25—27页。

W.)测得人在一瞬间只能辨认大约6个物体;1898年,厄德马和道奇(Erdma, B. & Dodge, R.)进一步发现,当英文字母随机排列并向被试呈现100毫秒时,被试只能报告出4～5个字母,他们采用的方法可称为"全部报告法"(whole report procedure),即要求被试看一张卡片上的字母刺激后,将其能够辨认出的字母全部报告出来。

斯珀林(Sperling, 1960)在重复上述"全部报告法"实验时,给被试呈现三行四列共12个英文字母,呈现时间为50毫秒,结果被试能报告的字母数也是4～5个。在斯珀林看来,这是一个非常简单的任务,为什么被试的成绩如此差,他认为一定存在一个更短暂的记忆系统,其对信息保持的时间很短,以至被试还没有来得及报告,有些信息就消退了,于是他设计了"部分报告法"(partial report procedure)实验:将一张写有三行四列共12个英文字母的卡片呈现给被试,呈现时间仍然为50毫秒,呈现完毕时随机给予被试一个声音信号。声音信号分高音、中音和低音三档,分别对应于卡片上的第一行、第二行和第三行字母。被试听到声音信号后,就立即报告出该信号对应的那一行的字母,其他行的字母就不用报告了,然后研究者用一行字母的报告率推算整体上可能报告的字母数。比如,被试看完卡片后听到了一个中音信号,他就报告第二行的字母。斯珀林以此方法缩短了报告时间,结果被试报告的成绩大大提高,比如被试能报告出一行字母中的3～4个,那么对于整张卡片来说,其可能报告的字母数就达到了9～12个。

斯珀林的实验结果表明:当呈现三行三列9个字母时,被试能根据不同的音调信号报告出一行字母,其正确率几乎达到100%;在呈现三行四列12个字母时,被试大约能正确报告一行字母的76%。由于声音信号是在字母刺激终止后出现的,被试只有在头脑中保持三行字母映象的情况下,才能根据音调信号报告出相应一行的字母。因此,按实验获得的百分比来推算,被试头脑中对字母保持的总量,在呈现9个字母时,应是9×100%=9个字母;在呈现12个字母时,应是12×76%=9.12个字母,都远远超过全部报告法实验中报告的字母数。

图8-10显示了斯珀林使用"全部报告法"和"部分报告法"对同一材料实验的结果。在全部报告法实验中,无论呈现多少项目,被试能正确报告的项目数平均为4.5个,而根据部分报告法实验计算出来的可报告项目数都超过全部报告法报告的项目数。斯珀林为了证明视觉信息在短时间内迅速消退,进一步改变实验安排:在刺激呈现完毕到提示信号出现之间插入不同的时间间距,迫使被试在不同的时间间距后用部分报告法重现项目,结果如图8-11所示。图8-11显示的是向被试呈现12个字母、提示信号延迟不同时距时的实验结果。结果显示,

被试报告的正确率随延缓时距的加大而迅速下降。当延迟时距达到 0.3～0.5 秒时，报告正确率下降速度减缓；达到 1 秒时，报告正确率已与全部报告法的结果基本一致。可见，“刺激刚刚消失时，比消失 1 秒后，图像记忆中保持的字母数要多得多。还说明，图像记忆的容量虽大，但消失得很快，接受的信息如不能从瞬时记忆转入短时记忆，很快就消失了……视觉的感觉记忆的作用时间似乎在 0.5 秒以内，约为 300 毫秒”。①

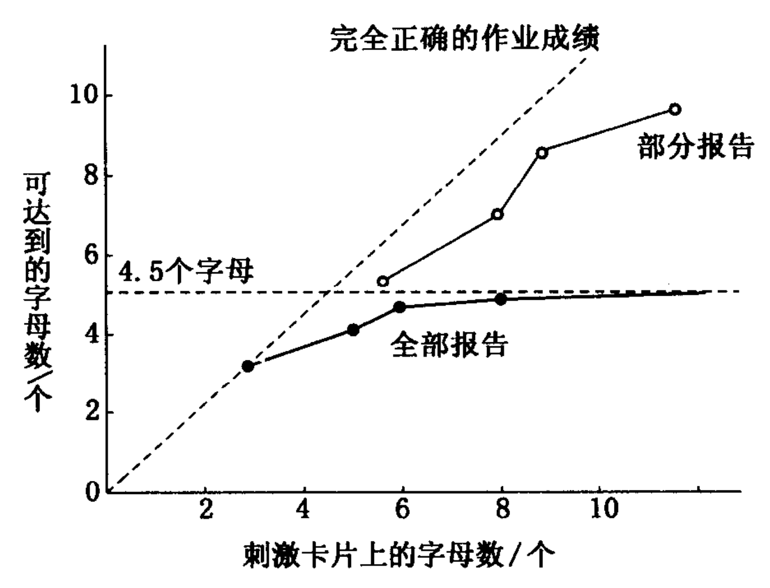


图 8-10　部分报告法和全部报告法的比较(Sperling, 1960)

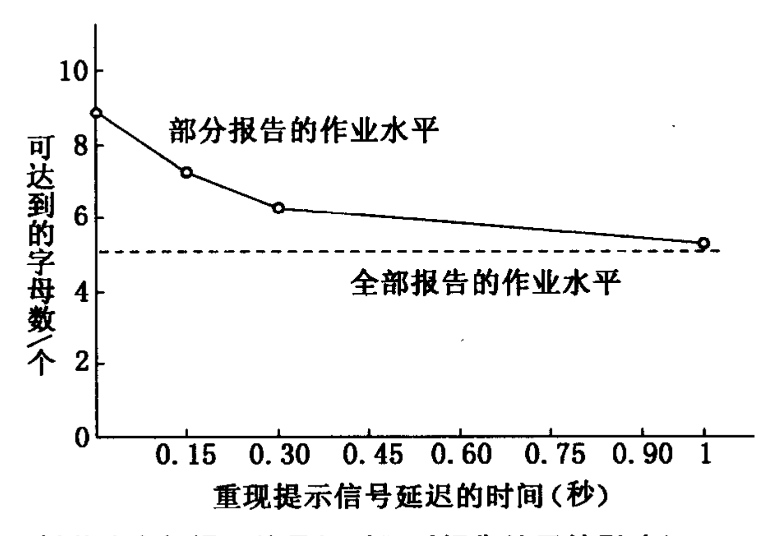


图 8-11　刺激消失与提示信号间时距对报告结果的影响(Sperling, 1960)

在斯珀林提出瞬时记忆及图像记忆概念之后，有研究者进行了一系列改进实验，进一步证实了感觉记忆的存在，基本上认为图像记忆保持信息的时间在

① 鲁忠义、杜建政：《记忆心理学》，人民教育出版社 2005 年版，第 47—48 页。

300 毫秒左右，容量至少在 9 个项目以上。

(二) 部分报告法实验与声像记忆

仿照斯珀林的视觉刺激部分报告法实验，达尔文等人(Darwin et al.，1972)开展了听觉刺激的部分报告法实验，即“三耳人实验”。他们给被试的双耳分别戴上一个立体声耳机，通过每个耳机的双声道，给被试左耳呈现两个刺激(如一个字母 B、一个数字 8)的同时，给右耳也呈现两个刺激(如一个字母 F、一个数字 8)，其中包含一个与左耳接受到的相同的刺激(如数字 8)。如果两耳接受到的这些刺激均在强度上平衡，那么被试就会分别在两耳以及两耳的“中间”(就像在“后脑勺”也有一个声音刺激)各听到一个声音信号(如分别为 B、8、F)，这就是所谓的“三耳人实验”。

左耳	“中间耳”	右耳
B	8	F
2	6	R
L	V	10

图 8-12 听觉刺激的全部或部分报告实验材料举例(Darwin et al.，1972)

利用“三耳人实验”，研究者给被试“三耳”分别呈现三个项目，如图 8-12 所示。达尔文等人也分别采用了全部报告法和部分报告法进行实验，其中部分报告实验的操作方法是：给被试呈现听觉刺激后，立即或延迟一段时间后在被试正前方屏幕的左、中或右的位置打出一个光条，提示被试报告与光条位置对应的“耳朵”听到的信息。实验中也安排了不同的提示信号延迟时距，分别为 0 秒、1 秒、2 秒、4 秒，实验结果也显示信息呈现后迅速衰退，如图 8-13 所示。不过，声

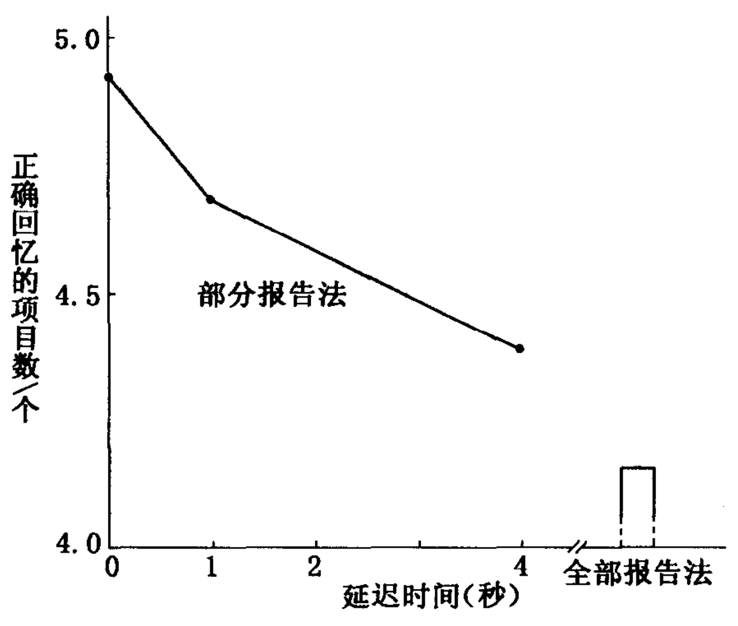


图 8-13 声像记忆的实验结果(Darwin et al.，1972)

像记忆与图像记忆具有一些明显不同的特征。首先,在声像记忆部分报告实验中,提示信号延迟大约 4 秒时,被试正确回忆的项目数才会接近全部报告实验的结果,这说明声像记忆对信息保持的时间在 4 秒左右;其次,声像记忆的信息容量要比图像记忆的容量小得多,在 5 个左右。

短时记忆和感觉记忆的证实推动了多存贮理论的发展。阿特金森和希夫林(Atkinson & Shiffrin, 1968)在上述研究基础上,提出了记忆的多存贮模型(multistore model of memory)或三级加工模型。该模型认为人的记忆系统由三个子系统构成,即瞬时记忆系统或叫感觉记忆系统、短时记忆系统和长时记忆系统。这三个记忆系统对信息储存的方式、表征方式、信息保持的时间、信息容量都不同。但这一理论也受到许多心理学家的质疑,甚至坚决反对,这其中以克雷克和洛克哈特(Craik & Lockhart, 1972)的批评最具代表性。

三、加工水平说及其实验证实

克雷克和洛克哈特认为把记忆看成由几个系统组合而成,这将记忆简单化了,同时也回避了许多更有意义的问题。因此,只有用“加工水平”(level of processing)的概念才能更好地阐明记忆问题。他们认为,外部信息进入加工系统后,要经历一系列不同水平的加工,从表浅的感觉分析,到较深的、更复杂的、抽象的语义分析,表浅的感觉分析主要涉及刺激的物理特征,而较深的分析则涉及刺激的意义或语义分析。记忆痕迹实际上是与不同的信息加工深度相联系,记忆痕迹深浅不同,信息保持时间的长短也就不同,记忆痕迹是信息加工的副产品。克雷克和图尔文(Craik & Tulving, 1975)用不随意学习(incidental learning)实验验证加工水平说的设想。实验中,他们把被试分成三组,三组被试分别针对呈现的 20 个单词回答不同的问题:

第一组被试回答该词是大写字母还是小写字母;

第二组被试回答该词是否与某词(weight)押韵;

第三组被试回答该词能否填入“他在街上遇见________”这一句子中。

显然,第一组被试只需要对刺激材料的书写特性进行加工,这是视觉物理特征水平上的加工;第二组被试只需要对刺激材料的语音特征进行加工,也是物理特征水平的加工;第三组被试则要在语义水平上对刺激材料进行加工。

完成这些任务后,突然宣布对所有被试进行再认测验,结果如图 8-14 所示。

从图 8-14 看出,对刺激材料的加工水平越深,反应的潜伏期越长,随后进行单词再认的成绩也越好。

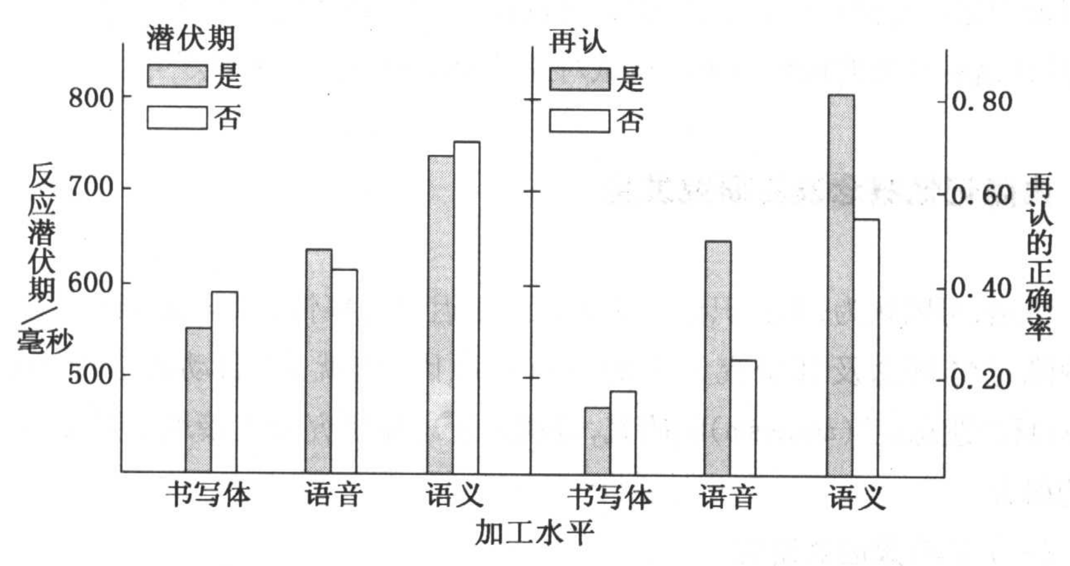


图 8-14 加工时间与再认成绩是加工深度的函数

(Craik & Tulving, 1975)

加工水平说虽然在解释不同深度的加工及其对信息保持的时间方面较为灵活,但还是招致不少批评。这些批评主要有三方面:①(1)加工水平说似乎是告诉人们,越是意义丰富的事件记得越是清楚和牢固,这其实是人们的常识,所以从这个角度说,加工水平说对人们理解记忆没有提供多少有价值的新见解;(2)加工水平说很含糊,没有给出加工深度的客观测试指标,所以该理论也难以得到直接验证;(3)它陷入循环论证的误区,一方面指出,加工越深,记忆痕迹就越强,信息保持的时间就越长;反过来又说,记忆信息保持时间越长,是由于对其进行了越深的加工。

第四节 内隐记忆的实验研究

心理学家普遍承认,人的心理过程的几乎所有方面都存在意识与无意识两方面,这也包括记忆过程。近年来,关于无意识的内隐记忆的实验研究得到了一系列有趣的结果,也给认知心理学带来另一片繁荣领地——内隐认知(implicit cognition)的研究。按照意识性划分,记忆可分为外显记忆和内隐记忆两大类。

① 鲁忠义、杜建政:《记忆心理学》,人民教育出版社 2005 年版,第 38 页。

外显记忆就是在意识指引下并能被记忆者本人觉知的记忆，前述的艾宾浩斯范式下的记忆研究基本上都属于外显记忆。“内隐记忆的根本特征是被试并非有意识地知道自己拥有这种记忆，但它可以在对特定任务的操作中自然地表现出来，并且这种任务的操作不依赖被试对先前经验的有意识恢复。”①

一、内隐记忆概念及其研究发生

虽然，可以认为，无意识记忆受到关注是迟早的事情，但毕竟有两个领域成为内隐记忆概念及其研究发生的更直接原因，这就是“启动效应”(priming effect)和“遗忘症”(amnesia)的研究，特别是遗忘症研究最直接地引发了内隐记忆的研究。

(一) 启动效应的研究

20 世纪 60 年代以来，认知心理学研究开始注意到一种奇特现象：人们经历过的事件或学习过的项目，有时不能进行有意识的回忆或再认，但却能在某些相关任务中间接表现出它依然存在的迹象，科弗(Cofer, 1967)将该现象称为启动效应。朱滢教授(1993)认为，启动效应指的是执行某一任务对后来执行同样或类似任务的促进作用；杨治良(1998)则认为启动效应是指由于近期与某一刺激接触而使对这一刺激的加工易化。不过，启动效应的最主要特征还在于它是发生在前后两种不同的刺激情境之间或两种不同的测试任务之间的，比如：先接受“学生”一词的刺激，后测量到被试对“教师”一词的反应时间缩短了，这是不同刺激项目间发生了启动效应；先抄写过一些单词，后让被试识别快速呈现的抄写过的和未抄写过的单词，被试对抄写过单词的识别率提高、反应时间缩短，这是前后两种测试任务不同而发生了启动效应。启动通常可分为直接启动(direct priming)和间接启动(indirect priming)两种。直接启动也叫做重复启动(repetition priming)，是指学习阶段呈现的刺激项目与测试阶段呈现的刺激项目完全相同；在间接启动(indirect priming)中，学习阶段与测试阶段呈现的刺激项目有关但不相同。

启动效应最常用的测验方法包括单词确定、单词辨认以及词干或词段补笔。在单词确定测验中，向被试呈现一个一个的字母串，要求被试说明每一字母串是否构成了一个合法单词，实验中记录被试的反应时间(Scarborough, Gerard &

① 郭秀艳：《实验心理学》，人民教育出版社 2004 年版，第 459 页。

Corstese, 1979)。如果某一字母串先后出现了两次,而且第二次出现时被试的反应时间减少了,则显示出启动效应。在单词辨认测验中,先快速呈现某一刺激,要求被试辨认,然后当被试再次遇到该项目时,辨识精度的提高或反应时间的下降则可作为启动效应的指标(Feustel, Shiffrin & Salasoo, 1983)。在词干或词段补笔测验中,研究者在前期的一项作业中向被试呈现过一系列词汇刺激,然后让被试参加补笔测验:向被试呈现一个词根或词段,要求他们尽快地用想起来的第一个合适的词补齐它。如果被试更多地使用了前期呈现过的单词,则反映了前次接触这一单词的启动效应(Graf, Squire & Mandler, 1984)。

(二)遗忘症的研究

20世纪60年代,神经心理学关于遗忘症的研究成为引发内隐记忆研究热潮的关键。1970年,英国神经心理学家韦斯克兰茨和沃林顿(Weiskrantz & Warrington, 1970)发表了关于遗忘症的系列研究。他们采用了一些残缺线索方法,发现了记忆过程中的分离现象。他们以4名遗忘症患者(3名科萨科夫综合症患者、1名额叶切除患者)为实验组,16名无脑损伤、记忆力正常者为对照组,让他们学习词表,然后进行不同类型的测验:回忆、再认、模糊字辨认、词干补笔。结果表明,严重的遗忘症患者在再认和回忆作业上均表现出明显的记忆障碍,但在另外两项测验中却与对照组没有明显差异。随后的一系列研究证实,遗忘症患者仍保留有学习和记忆能力。这种能力仅对部分线索法或间接测量方法敏感,传统的直接测量方法测量不出来。①于是,研究者认为,传统的直接方法和部分线索间接方法测量的是两种不同的记忆,格拉夫和沙克特(Graf & Schacter, 1985)将其分别称为外显记忆和内隐记忆。

可见,遗忘症患者的外显记忆受到了破坏,而内隐记忆并未受到明显的影响,暗示了内隐记忆是独立于外显记忆系统的另一个记忆系统。那么,为什么会出现这种记忆的分离呢?不同的理论给出了不同的解释。

二、对内隐记忆的两种解释

内隐记忆的内在机制是什么?目前有两种代表性的理论解释:一是图尔文和沙克特(Tulving & Schacter, 1990)提出的多重记忆系统理论,一是罗迪格等

① 鲁忠义、杜建政:《记忆心理学》,人民教育出版社2005年版,第231页。

人(Roediger et al., 1989)提出的对应加工迁移理论。

(一) 多重记忆系统理论

多重记忆系统理论(multiple memory system theory)“将记忆看成是由多个服务于不同对象目标、遵从于不同操作原则的子系统共同作用而实现的一种功能”①,其中每一子系统又由若干特定的加工过程来完成。“同一个子系统内的加工过程之间的关系比不同子系统中的加工过程间的关系更加密切,并且在理论上每个记忆系统都可能具有其特定的神经机制与行为指标。”②看来,外显记忆和内隐记忆可被看作是两个相对独立的记忆子系统。

对多重记忆系统理论合理性的阐述和证明是围绕着外显记忆与内隐记忆是否具有结构机能独立性而展开的(Tulving, 1985; Witherspoon & Moscovitch, 1989)。图尔文认为,内隐记忆代表了一种新的记忆系统,即知觉表征系统(perceptual representation system)。知觉表征系统的信息储存是分布的,它缺乏某个中心的痕迹。当某一不完整的情景线索激活对应的部分信息时,个体就可以利用这些线索激活部分信息并完成某项任务,但由于没有指向中心的记忆痕迹,所以它不实现对整体事物信息在头脑中的重现。图尔文等人(1990)为证明其理论的合理性,开展了进一步的实验研究。结果表明,在外显记忆测验中,针对同一目标的两种不同线索呈中等程度的相关;而在内隐记忆测验中,针对同一目标的两种不同线索则是相互独立的。由此证明内隐记忆和外显记忆是以不同的结构或机制表征信息的,内隐记忆是“无痕迹”的记忆系统(traceless memory system)。

多重记忆系统理论能较好地解释遗忘症患者的记忆分离现象。因为各个记忆系统是相互独立的,所以当其中某一记忆系统受损时,另一记忆系统依然可以正常工作。比如,当陈述记忆系统受损伤时,程序记忆系统能保持完好。但是,记忆正常的人也会出现记忆分离现象,这一点就难以使用多重记忆系统理论来解释了。此外,多重记忆系统理论还有一个无法回避的、至今都未能很好解决的根本性问题:多重记忆系统的划分标准是什么。

(二) 对应加工迁移理论

罗迪格和布兰克斯顿(Roediger & Blaxton, 1990)不认为内隐记忆和外显记忆是属于不同的记忆系统,而主张把它们看成是单一记忆系统的不同方面,为

---

① 鲁忠义、杜建政:《记忆心理学》,人民教育出版社 2005 年版,第 265 页。

② 郭秀艳:《实验心理学》,人民教育出版社 2004 年版,第 461 页。

对这单一记忆系统进行解释，他们提出了对应加工迁移理论（transfer-appropriate procedures approach）。按照这一理论，记忆中的实验性分离现象，只是由于不同测验任务所要求的加工过程不同造成的，并不是由于受相互独立的两个不同记忆系统控制所致。具体地说，如果记忆测验所要求的加工过程正好与学习时的编码加工一致，或者基本一致，测验的成绩就会比较好，否则就会比较差。罗迪格等主张按照所要求的心理加工过程来划分不同的记忆测验，即：是要求概念驱动的过程，还是要求数据驱动的过程。他们认为外显记忆测验要求的主要是概念驱动的过程，即要作有意义加工、精细编码和心理映象等加工过程，正是由于对材料的精细编码以及有意义的组织，外显记忆测验的成绩才得以提高。相反，多数内隐记忆测验在相当程度上是依赖学习时的知觉过程与测验时的知觉过程的对应程度，并几乎都是提取过去经验中的知觉成分，因而可以认为内隐记忆测验要求的主要是材料驱动的过程。所以，影响概念加工的各种变量对内隐记忆就没有影响，而行为特征方面的变量对内隐记忆就会有很大影响；反之亦然。①

对应加工迁移理论可以对记忆正常的被试表现出来的直接和间接测验间的分离现象提供较好解释，但却难以解释遗忘症患者的记忆分离现象。按照对应加工迁移理论，遗忘症患者外显记忆效果较差而内隐记忆效果较好，是因为其概念驱动过程有障碍而材料驱动过程保持完好。但是，在有些要求有概念驱动加工的间接测验中，遗忘症患者却能表现出与正常人同样的水准，对应加工迁移理论难以对此给出说明。

显然，上述两种理论一方面都可以解释一些记忆分离现象，但都不能解释所有的记忆分离现象，而且在关于内隐记忆内部机制的解释上还带有许多猜想的成分。不过，我们也看到，两种理论在某些方面并非完全对立，比如，多重记忆系统理论认为内隐记忆主要代表知觉表征系统，而对应加工迁移理论则提出内隐记忆主要是依靠数据驱动的加工过程，在这一点上二者似乎“不谋而合”。所以，目前有一些研究者，正试图将两种理论结合以期能有效解释各种记忆分离现象。

三、内隐记忆研究的实验方法

“内隐记忆实验的首要任务，在于将内隐记忆和外显记忆区分开来，在此基

① 郭秀艳：《实验心理学》，人民教育出版社 2004 年版，第 462—463 页。

础上才有可能建构内隐记忆和外显记忆两者不同的因变量指标,并进而考察自变量对内隐、外显记忆的不同影响。因此,如何在实验中分离内隐记忆和外显记忆,就成为内隐记忆实验研究的逻辑重点。”①当前,在内隐记忆实验研究中使用的分离方法主要包括两大类:一类叫做实验性分离(experimental dissociation),也叫做任务分离(task dissociation);另一类叫做加工分离(process dissociation)。

(一) 实验性分离的逻辑与方法

实验性分离范式最早为神经心理学家所采用,主要用以研究脑损伤病人心理功能的分离现象(Teuber, 1955)。当前,实验性分离已经成为认知心理学的重要方法之一,杨治良(1996)甚至称其为实验心理学的三大新方法之一。实验性分离的逻辑认为,我们通过一系列不同的测验而取得被试心理活动的因变量,其中一些因变量对应于内隐记忆,另一些因变量则对应于外显记忆,因此实现记忆系统的实验分离。这种分离主要是通过直接的和间接的不同测验完成的。

直接测验就是在测验指示语中明确要求被试有意识地回想他们经历过的某些事件并把它们从记忆中提取出来,如自由回忆、再认测验等,在这些测验中被试直接从意识中提取信息。间接测验则是在测验指导语中并不要求被试有意识地提取过去学习的信息,而是通过他们在一些特定任务上的表现来间接推断被试是否对某些信息拥有记忆。研究者认为,这两类任务分别对应于外显记忆和内隐记忆,因而这种分离也被称为任务分离。这里包含的逻辑是:如果这两类测验的成绩表现出同样的变化趋势,或者说具有高度的正相关,那么这两类测验中信息的提取依赖同样的心理结构;如果两类测验的成绩表现出完全不同的变化趋势,甚至相反的趋势,则说明二者依赖不同的心理结构。图尔文曾对实验性分离作如下表述:“符合实验性分离逻辑的实验是这样的:控制单一的变量而比较在两种不同的任务中变量的效应……如果变量影响被试在一种任务中的操作,但不影响另一种任务的操作,或者变量对两种任务的操作的影响有不同的方向,我们就说分离产生了。”②

在内隐记忆研究领域,实验性分离实验中的直接测验就是指前述的回忆、再认等直接信息提取方法,间接测验的主要方法是补笔测验和知觉辨认。

---

① 郭秀艳:《实验心理学》,人民教育出版社 2004 年版,第 463 页。

② Tulving, E. (1985). How many memory system are there? *American Psychologist*, 40: 385—398.

1. **补笔测验**

补笔测验主要包括词干补笔(偏旁补笔)、残词补笔、残图补全等。词干补笔的实验程序是[1]:首先要求被试学习一系列单词,而后主试向被试呈现每个单词的前三个字母,并要求被试用想到的单词填写出来。主试提供的词干可填写成多个有意义的单词。结果被试会更多地使用学习阶段呈现过的单词完成补笔测验,被试用呈现单词补笔的概率减去用未呈现单词补笔的概率,就是他所学单词的启动量。残词补笔和残图补全的实验程序与词干补笔基本相同。

虽然补笔测验中使用的都是一些长时记忆信息或知识,但研究发现,这些知识的提取也依赖实验前期发生过的事件。例如,在完成 ele ______时,如果被试最近学过"elephant"这个词,尽管可以有多种完成的方法(如 element, electric, elephant),但他们有更大可能会用"elephant"这个词来完成词干补笔或残词补全。通过在这些测验中表现出来的启动量就可以衡量内隐记忆的效应。[2]

2. **知觉辨认**

知觉辨认法的测验程序是:在学习阶段,向被试依次呈现一系列项目,然后将这些呈现过的项目与未呈现过的项目混淆在一起,在速示仪或电脑显示器屏幕上以极短的时间(小于 5 毫秒)一个一个地呈现给被试,要求被试加以辨认。一般来说,被试将学习阶段呈现过的项目辨认出的概率高于未呈现项目,二者的概率之差就是呈现项目的启动量。知觉辨认的材料可以是文字的,也可以不是文字的,如单词、伪词、无意义音节、图画、乐曲等。知觉辨认的另一种变式为模糊字辨认,是指在测验时呈现单字的字母很模糊,要求被试将它们辨认出来。

雅各比(Jacoby, 1983)采用再认和知觉辨认两种方法进行了实验。他在三种条件下让被试学习反义词:无上下文条件,被试看到"×××——冷"便大声说出"冷"字,这里要求被试加工的是"冷"字的字形;有上下文条件,被试看到"热——冷"并大声读出"冷",在这里是要求被试加工"冷"字的语义;产生(generation)条件,被试看到"热——?",而要求被试说出"冷"字,这是前两种条件的结合,要求被试既要加工字形,也要加工语义。在这三种条件下,他均将被试随机分成两组,一组被试使用直接测量法进行测验,即再认;另一组被试则用间接测量法进行测验,即知觉辨认。结果发现,采用直接测量的一组被试在产生条件

---

① 鲁忠义、杜建政:《记忆心理学》,人民教育出版社 2005 年版,第 233 页。
② 郭秀艳:《实验心理学》,人民教育出版社 2004 年版,第 465 页。

下的击中概率最高,有上下文条件次之,无上下文条件最差;而采用间接测量的一组被试所得结果正好相反:无上下文条件下击中概率最高,有上下文条件次之,在产生条件下最差。从这一实验,研究者有效地观察到了实验性分离的记忆结果。

(二) 加工分离的逻辑与方法

实验性分离实验为内隐记忆积累了大量证据,也有效地揭示了外显记忆和内隐记忆的不同机能特点,但它也还存在诸多问题,受到一些学者的批评。批评的意见主要有以下两点。

首先,直接测验和间接测验的可比性问题。实验性分离的逻辑是利用不同测验结果的不一致性证明二者依赖不同的记忆系统,进而证明外显记忆与内隐记忆是不同的系统。这里的问题是,即使当外显记忆与内隐记忆不是相互分离的两个系统,我们用不同的任务分别去测量它们,结果也还是可以出现不一致的趋势,因为测验任务本身就存在测验线索、记忆任务、反应指标等方面的差异。其实,从某种意义上说,二者本来就没有可比性。

其次,在内隐记忆是否存在的证明上存在循环论证之嫌。一方面,研究者假设,直接测量对应于外显记忆,间接测验对应于内隐记忆,然后再来使用这两种测量结果的差异证明内隐记忆的存在。那么,如果两种测量并不是各自纯粹地对应于两种记忆,而是两种记忆在两种测量任务中交叉存在,只是程度不同而已,前述的假设也就难以成立了。

看来,使用任务分离技术来证明和测量内隐记忆,还存在不少疑点。于是,雅各比等人在 20 世纪 90 年代又提出了加工分离程序(process dissociation procedure,简称 PDP),以试图解决上述方法学的困难。

典型的再认型的加工分离程序是这样的①:在学习阶段先后给被试呈现两个词表,然后将一些未学习过的新词与这些词混在一起,进行再认测验。在测验阶段,分"包含"测验和"排除"测验两种。如果被试进行包含测验,则要求被试要将学习过的两个词表上的项目判断为"旧"的项目(即判断为"学过"),把后来混进去的项目判断为"新"的项目(即判断为"未学过");如果被试进行排除测验,则要求其只是将第二个词表中的项目判断为"旧"的项目(即判断为"学过"),把第一个词表上的词及后来混进去的词判断为"新"的项目(即判断为"未学过")。就

① 鲁忠义、杜建政:《记忆心理学》,人民教育出版社 2005 年版,第 252—253 页。

第一个词表来说,有意识的控制加工和无意识的自动加工在包含测验的作用是一样的,即有意识的回忆和熟悉感都会让被试作出"旧"的反应,两种加工在测验中的作用是一致的。在排除测验中,意识与无意识的作用相反,如果被试能有意识地回忆其属于第一个词表,就会作出"新"的反应;如果被试不能回忆但感到熟悉,就会将这个项目判断为第二个词表项目,而作出"旧"的反应(Jacoby,1991)。由此可见,这一程序实现了自动加工与控制加工的分离。

加工分离程序的逻辑依赖雅各比提出的三个理论假设:"(1)熟悉感对一组特定项目的影响在包含和排除测验中是相同的;(2)有意识回忆对两个测验操作的贡献程度也是一样的,这两个假设可归纳成 PDP 的'一致性假设';(3)意识加工和无意识加工是彼此独立的,它们之间呈零相关。"①说得明白些,无意识信息提取导致的熟悉感,在包含测验和排除测验中都会使被试作出相同概率"旧"的反应("学过的");有意识回忆的项目会使被试在包含测验中将其报告为"旧"的,而在排除测验中,会让被试将其报告为"新"的,这两种事件的发生都是由于被试能够有意识地回忆起这一项目。

在上述加工分离程序中提供了两种测试条件:"包含"条件,意识成分和无意识成分共同促进作业成绩,它们的关系是协同的;"排除"条件,意识成分和无意识成分对作业成绩的影响正好相反,它们的关系是对抗的,从而实现了两种信息提取过程的分离,并能通过具体的数据计算,将混合于各种任务之中的意识成分和无意识成分分离出来。"加工分离程序在体现了对抗逻辑的思想的同时,为分离自动的和意识控制的记忆加工提供了一个有效的途径。在对这个创造性构想不断进行修正和批评的过程中,内隐记忆的方法论和实验技术得以进一步完善和发展,并使早日揭示人类记忆的本质成为可能。"②

阅读材料 8-1

脑结构与记忆

学习记忆是极有临床和实际应用价值的研究课题。然而,由于学习记忆过

① 鲁忠义、杜建政:《记忆心理学》,人民教育出版社 2005 年版,第 253 页。
② 郭秀艳:《实验心理学》,人民教育出版社 2004 年版,第 469 页。

程十分复杂，我们至今对它的了解仍然很有限。以往的研究曾经发现，在某些简单学习记忆行为方面，诸如习惯化、敏感化及经典条件反射形成，低等动物与人类遵循共同的行为原则，神经细胞电活动规律也是相似的。随着动物的进化及其生存环境复杂性的相应增加，神经系统逐渐形成了许多相对特殊化的结构，用以完成特定的功能。在高等动物神经系统中，海马被认为是短时记忆形成的关键结构。当海马受损伤时，近期记忆不能保存，而远期记忆不受影响。根据这一特点，临床上对一些疾病的定位或早期诊断有很大帮助，如脑出血、颅脑创伤、颅内肿瘤及早老性痴呆等。目前的研究认为，海马的作用类似一个加工平台，仅接受信息予以暂时保存，然后将信息分送到其他脑区长期保存，形成以蛋白质合成增加为特征的长时记忆储存。长时记忆还涉及神经回路、突触数量及突触效率的变化。人类的记忆行为则更加复杂，例如，根据信息传入途径不同，可分为视觉、听觉、触觉、嗅觉及动觉记忆；根据信息内容不同，可分为语言、图像、时间、空间及情感记忆等。就人类大脑而言，目前对语言记忆中枢有相对较多的了解，如中央前回皮层布罗卡(Broca)区、枕叶锯状裂附近皮层及颈叶角回皮层分别与运动、视觉、听觉语言记忆有关。研究者最感兴趣的问题仍然是学习记忆功能所依附的神经区域，以及各种神经区域之间如何协同作用以完成对信息的加工、整合、储存及提取(隋南，1999)。

空间知觉由距离、深度及方向等三种知觉构成。我们的研究工作已经证实了距离、深度及方向三种知觉在行为方式上的可分离性(隋南等，1992，1993，1995)。在莫里斯迷宫中，海马损毁大鼠寻找隐匿站台是根据自身位置与迷宫边缘之间的距离来判断站台位置的，表现出方向知觉和深度知觉障碍，但距离知觉仍然存在。这种“边缘式”(paratic)策略是以自我参照点(egocentric)的认知方式，与正常大鼠采用的以外部线索为参照点(allocentric)的“图式”(mapping)策略有本质的区别。大脑皮层(如顶叶或额叶)损毁大鼠采用外部线索为参照点的“趋向式”(taxic)策略搜索目标，但是方向知觉和深度知觉显著弱化。而隔区损毁大鼠的“随机式”策略则同时表现出方向、深度和距离知觉的障碍。研究结果提示：异我参照点的认识加工主要是在大脑皮层以下特别是海马等结构完成；而自我参照点的认知方式则属于趋于上移或进化上发生较晚的功能，主要由大脑若干部位协调完成(隋南等，1992，1993，1995，1997)。隔区则在空间认知整合过程的不可替代性亦被证实(隋南等，1992)。利用损毁和药理学方法对雏鸡一系列脑区的研究发现，雏鸡的中间内侧上纹状体腹核(IMHV)和旁嗅球叶

(LPO)是参与学习记忆活动的重要脑区(胡家芬、隋南等,1999,2000;高阳、匡培梓等,1997,1998)。通过有关即刻早基因(c-fos,c-jun)表达的研究也证实了这一结论(高阳、匡培梓等,1997,1998)。更深入的研究则表明IMHV与学习记忆的贮存有关而LPO与学习记忆的巩固关系更为密切。阐明认知加工的不同层次及不同层次加工系统所依附的神经结构,进一步了解学习记忆相关递质、激素和肽类对学习记忆的影响及其机制,对揭示脑活动规律具有重要作用。

——资料来源:《脑、激素与记忆》(隋南,2001)①

阅读材料8-2

记忆研究的应用

对人类记忆的研究已应用于很多情境之中……现代生活中对记忆的需求在不断增加。复杂的电话系统、计算机系统、传真系统、照相机、录像机、汽车等等引起了不断增加的认知需求,常常是在记忆方面。仅仅需要记住的密码数和电话号码就超出了很多人的记忆容量。人们常常要求心理学家用其专长来解决这些实际的问题。特别是电话公司和计算机公司雇用了很多心理学家进行工作,通过考虑人的因素使装备发挥作用,这些工程心理学家,正如他们被称呼的那样,试图设计复杂的系统,以便使人类操作者在使用过程中相对地少犯错误。例如,正如我们前面所讨论过的,英国和美国电话公司所赞助的早期研究表明,在立即测验中能够被记住且不犯错误的电话号码的最大位数是7,所以电话号码被做成7位数。这样,这些心理学家证明如果形成组群或组块,这些数字将能够得到更好的回忆,所以美国的电话号码典型地组成三个、然后是四个一组,如792-3948。在法国,电话号码是8位数,他们以四个两位数的方式呈现:79-23-94-81。虽然工程心理学家们还必须考虑记忆之外的其他过程,但是由工程师制造的复杂的机器装置必须考虑它们的使用者工作记忆的容量限度。心理学家们经常帮助设计这种人-机界面。

教育心理学家也应用人类记忆的研究成果。毕竟,教育中的一个主要话题

① 隋南:《脑、激素与记忆》,见中国心理学会:《当代中国心理学》,人民教育出版社2001年版,第124—129页。

(也许就是中心话题)是将知识从老师和书本传输给学生,以便学生能够学习和记住这些信息。因此,心理学家用了很多时间来研究如何在教科书中组织和呈现信息,如何讲授这类话题。一个问题是如何将从人类记忆基础研究获得的信息传输给可能会使用它的教育者们。有很多心理学家的发现已被应用到教育实践中。

心理学家还曾设计认知康复的程序。例如在很多医学条件下,如阿尔采默氏病(早老性痴呆,Alzheimer's disease)患者的一般认知功能,特别是记忆功能损伤。认知心理学家已设计出程序来改良(但是肯定不是治愈)人们在回忆中遇到的严重困难。

与此有关的是,很多心理学家也在研究增进记忆的技术,并且已经开发了一些课程,写了一些书为增进记忆提供一般的建议。很多记忆术自古已广为人知,但是现代认知心理学家已研究了这些系统所提供建议的特定方面,看看这些技巧中哪些特征确实有用。例如,如果将表象用作记忆的手段,那么为了有效,把一些表象想成是古怪的,有必要吗?正如上面讨论过的,答案并不是很清晰,但在某些特别困难的项目中有选择地使用古怪的表象应当会提高成绩。

研究记忆的认知心理学家还经常被作为专家受到邀请,以处理需要依靠人类记忆变化规律来解决的法律问题。在此,我们将提到三种案例。首先,很多案件的关键需要亲眼所见的鉴定。如果在一次犯罪中有证人证实一个嫌疑人就是凶犯,这一事实代表法庭上的有力证据。然而,目证的准确性如何呢?对此问题有很多研究,发现证人有时可以是错误的,尽管十分有信心地说自己是对的。已发现促成目证错误的条件,心理学家在司法案例上验证了这项研究。

第二种庭审案例涉及儿童作为证人。儿童可以作为可靠的证人吗?他们的回忆有多准确?什么情况下他们可能被误导作出错误的控告?这些问题在当儿童作为主要证人或唯一证人时是非常重要的。因此,研究记忆发展的心理学家就是这些案件中的专家。

第三种庭审案例的情况涉及经过长时间后的记忆恢复。近年在美国、加拿大、英国均有很多这方面的新闻。有一个典型的例子是一位30或40多岁的妇女由于抑郁或某些其他问题进行心理治疗。在一次正常治疗过程中,治疗者可能问到她的童年经历。在谈论她的童年时,或者在回答治疗者的询问或请求时,这个病人可能回忆起从那个阶段开始的一些不愉快事件。在最为戏剧化的例子中,她可能回忆起曾经受到她父亲,或一个叔叔或一个老师的身体或性方面的虐

待。在一些案例中，这些在治疗过程中恢复的记忆第一次暗示了病人确实有过这样的经历。一种解释是所经历的这些事件是如此可怕，因而被压抑起来或者被丢弃到无意识状态中，然后在治疗过程中，这些记忆恢复了。另一种解释是这些事件根本就没有发生过，是在治疗过程中通过治疗者的暗示或其他机制创造了这些"记忆"(Lindsay & Lead, 1984)。

在这种记忆恢复的这些案例中的问题是非常复杂的，当这些问题在法庭上提出时，是向司法系统提出挑战。心理学家作为专家证人作证。虽然对这些恢复记忆陈述的真实性有不同意见，但很多研究者都相信，对被虐待的记忆的恢复只有在还有其他可以证明的证据时(医院记录、其他证人等)，才能够被接受。这种记忆现象是著名的，特别是在经历了长时间的间隔以后；而很多证据指出记忆的暗示性是这些记忆有力的源泉。

——资料来源：《记忆过程》(Roediger & Meade, 2000)①

建议阅读文献

1. 杨治良等：《记忆心理学》，华东师范大学出版社 1999 年版，第 201—229、257—319、395—405 页。
2. 杨治良：《基础实验心理学》，甘肃人民出版社 1988 年版，第 285—300 页。
3. 鲁忠义、杜建政：《记忆心理学》，人民教育出版社 2005 年版，第 2—18、304—342 页。

复习思考题

1. 如何理解回忆法、再认法、节省法、序列重建法、提示法、对偶联合法、系列位置效应、近因效应、首因效应、两种记忆说、部分报告实验、图像记忆、声像记忆、启动效应、重复启动、间接启动、外显记忆、内隐记忆、实验性分离、加工分离、词干补笔、知觉辨认?
2. 艾宾浩斯在实验心理学史上的贡献主要有哪些?
3. 巴特利特的记忆研究与艾宾浩斯的研究有何不同? 他在记忆研究中的主要

① 张厚粲主译：《国际心理学手册》，华东师范大学出版社 2002 年版，第 180—182 页。

贡献有哪些？

4. 记忆实验研究的基本方法有哪些？
5. 如何通过自由回忆实验证实短时记忆？
6. 斯珀林是如何通过实验方法证实感觉记忆的？
7. 何谓加工水平说？它的实验依据是什么？
8. 启动效应和遗忘症的研究与内隐记忆研究有什么联系？
9. 关于内隐记忆的主要理论解释有哪些？
10. 实验性分离和加工分离有什么异同？它们是如何证实内隐记忆的？

第九章
思维的实验研究

本章内容提要

最早对思维进行实验研究的是符兹堡学派，该学派对重量判断、观念联想和思考问题都做过实验研究，提出了“无意象思维”的概念。行为主义和完形主义心理学在关于思维的研究中提出过“边缘思维论”、“尝试-错误说”和“顿悟说”，都产生过重要影响。当代著名认知心理学家纽厄尔、西蒙和肖的合作则开辟了思维实验研究的新局面，他们提出和运用了思维研究的计算机模拟、口语报告分析方法。在纽厄尔和西蒙看来，问题解决就是问题解决者对问题空间进行启发式搜索的过程，而问题空间就是问题解决者对问题的心理表征，它体现了问题解决者所发现的问题中介状态，以及实现这些状态转换的算子。把这些中介状态和操作算子按时间顺序排列，就构成了问题行为图。此外，本章还介绍了概念形成的实验过程、问题解决者的策略与方法等。

思维(thinking)是对事物本质属性和必然联系的反映，这种反映能够超越人类感官的局限性，认识那些无法通过感官直接感知的事物的属性，所以说它是高级的认知过程。由于这种高级心理过程的复杂性，传统实验心理学对之进行的研究很少。20 世纪 50 年代后，随着信息加工心理学的兴起，思维等复杂心理过程的研究有了新方法和新途径，更由于人工智能(artifical intelligence)开发和知识社会对人力资源的特殊需求促进了这一领域的发展。本章介绍思维研究的一般问题、特殊实验方法和经典实验，以展示这一领域研究的复杂性、可能性和可资利用的研究策略与方法。

第一节　思维研究简史

逻辑学(logic)和心理学都在对思维进行系统研究，但是二者研究的目的和

方法存在很大差异。逻辑学主要研究人们应该如何推理,如何推理才能保证得到真命题;心理学主要研究人们是怎样进行思维的,以及这些思维活动如何受各种内外因素的影响,为什么会具有创造性等。近年来,由于认知心理学内部发生的重大变化,思维研究表现出许多新趋势,如生态现实元理论强调社会文化背景和内容相关性对思维的影响,同时也在加紧研究思维的脑生理机制。应该说,思维是未来一个时期相当繁荣的研究领域。那么,思维心理学的研究经历了一个什么样的历史过程呢?从中我们又能得到什么样的启示呢?

我们可以把思维的心理学研究划分为四个阶段:哲学心理学的研究、早期实验研究、行为主义和完形主义的研究、信息加工心理学的研究。

一、哲学心理学的研究

思维研究可以追溯到古希腊的联想概念和联想定律,因为英国哲学家正是在继承了这些早期哲学思想之后建立了影响深远的联想主义心理学。联想主义心理学(associationistic psychology)对思维进行了开创性的研究,提出了一系列颇有见地的观点。①

洛克(John Locke, 1632—1704)说:“……我认为这个词(指观念)可以代表人们思维时理智的任何对象……我们的某些观念彼此具有一种天然的关系和联系,把它们结合在一起,乃是我们理智的职责。”在洛克看来,思维的任务就是要找到事物间固有的关系或联系,并依此将不同的事物结合在一起。贝克莱(George Berkeley, 1685—1753)因此说:“例如,某种颜色、滋味、气味、形象和硬度,如果常在一块出现,我们便会把这些观念当作一个单独的事物来看待,并用苹果的名称来表示它。”这里描述的是概念形成的过程和基础。他还说:“显然,当心灵并非直接感知任何观念本身时,他必然要借助其他观念……我们从一个人的脸色转变为潮红或苍白,就常常会看出他所表现出的羞愧和恐惧。”这里所指便是思维的间接性了。

休谟(David Hume, 1711—1776)说:“看来事物间的‘必然联系’这个概念,乃是由于这些事件在许许多多类似的事例中经常结合在一起而产生的……”显然,这是思维的概括性。还有哈特莱(David Hartley, 1705—1757):“任何感觉 A、B、

① 邵志芳:《思维心理学》,华东师范大学出版社 2001 年版,第 15 页。

C等等，经过足够次数的相互联系(即联想)，就会获得支配相应的观念a、b、c等等的力量，这样，这些感觉中的任何一个，例如A，即使单独发生，也一定会在心灵中引起a、b、c等其余的观念的。"这段话也说明了思维的间接性，并且包含图式思想。可见，联想主义心理学对思维的概括性、间接性和对概念形成过程都有了一定程度的认识，但这些认识并非出自系统的实验探索，多属哲学思辨的产物。

二、思维的早期实验研究[①]

冯特在1879年建立心理学实验室和实验心理学比较完整的体系时，同时提出了系统的自我观察法即内省法作为心理实验的方法，但是他同时指出思维等高级复杂的心理过程是飘忽不定、无法内省的，思维被逐出了实验研究的范围。冯特的学生屈尔佩(Oswald Kulpe, 1862—1915)则受到艾宾浩斯的启发并超越了导师的观点，最先真正把思维当作心理学专门研究的课题。他认为，艾宾浩斯可以用实验方法来研究记忆这一高级复杂的心理现象，思维过程当然也可以用实验方法进行研究，于是他与他的学生在德国的符兹堡大学对思维进行了大量的实验研究，其中具有代表性的如马尔比的判断研究、瓦特的联想研究、彪勒的思考研究，从而创立了心理学史上有名的符兹堡学派(Würzburg school)。

马尔比(Karl Marbe, 1869—1953)对判断进行实验研究。他要求被试先后举起两个物体，判断孰轻孰重(1901)。结果显示，被试的判断通常是正确的，但被试却无法描述判断是如何进入意识的，故马尔比认为判断不是一种意识过程，因为被试在实验中不能告知任何内容。[②]在此之前人们认为，比较轻重时，人们会保持住对第一个物体轻重的感觉与意象，而将其与第二个物体的感觉与意象进行比较。但是马尔比的研究表明，被试在比较轻重的时候脑子里好像是空的，那些感觉和意象并不存在。这就是说，在判断时，思维不能表现为感觉和意象。其实按照现代认知心理学的解释，马尔比研究的正是知觉过程中的判断，只不过这种判断是非常短促和压缩的，被试还难以意识到，所以"脑子里空空的"。

瓦特(Henry Jackson Watt, 1879—1925)的联想研究(1904)。实验时，要求被试作有控制的联想，例如进行种属联想(梧桐——植物)或整体部分联想(房

① 邵志芳:《思维心理学》,华东师范大学出版社2001年版,第16—17页。
② 林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社2003年版,第792页。

间——门)。瓦特把联想过程分为预备、呈现、寻找反应词、反应词出现四个阶段，同时要求被试进行内省和描述。研究者预期,被试的内省可以反映出寻找反应词和反应词出现时的心理条件。但结果发现,被试一接受到刺激词就立即反应,因此他们无法对上述各个阶段进行内省。瓦特认为,决定反应的主要是预备阶段——明确课题方向、确定方向。客观而论,瓦特提出了有意义的实验课题,却没有得到明确的有价值的研究结果,原因是什么呢？当时的实验技术尚达不到分析快速完成的心理过程的要求。这一缺憾由后来认知心理学中的反应时间分析技术所弥补。

彪勒(Karl Bühler, 1879—1963)的思考研究。研究中,主试提出若干需要思考方能解答的简单问题由被试解答,时间为5～20秒。回答后询问并记录被试得到答案的步骤。结果被试并不能讲清他们是如何得到答案的。彪勒在研究中没有得到多少被试的内省材料,这可能是因为他的实验材料属于简单问题,被试的思考过程非常短促、达到自动化程度,所以意识不到思考的中间过程。彪勒在与屈尔佩的共同研究中,于1908年发表3篇论文,总题目是《思维过程心理学的事实与问题》,提出思维不能归结为感觉和意象,而存在一种非感觉的、非意象的思维元素,进一步证明存在"无意象思维"(imageless thinking)的观点。彪勒还用提问法来弥补内省法的不足。①

符兹堡学派对思维进行了许多研究,但是几乎没有得到什么确切的研究结果,所以,屈尔佩离开符兹堡后,符兹堡学派也随之解散。其实,我们看到符兹堡学派提出的研究课题和研究思路都是非常有价值的,现在的许多研究也还是与此很相似。没有得到预期结果的原因相当程度上有研究技术的问题。因此,我们是否可以说,在那个时代,对思维这一高级、复杂、内隐的心理现象进行实验研究的时机还不成熟呢?

三、行为主义和完形主义的研究

(一) 行为主义的研究

行为主义一贯主张研究看得见、测得出的行为表现,为什么会研究思维呢?因为行为主义的创始人华生对思维持有一种"边缘思维论"(marginal thinking theory),即认为思维是作为一种整体的躯体的机能,所以在他看来,思维是可以

① 林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社2003年版,第71页。

观察得到的,是可以进行实验研究的行为内容之一。

华生虽然承认思维具有内隐性,但是他认为思维与语言有密切关系,它在很大程度上受语言机制的影响。他说,思想,只是自己对自己说话。这样,心理学就可以通过言语这种外显的行为来研究思维了。"出声言语中所习得的肌肉习惯,也负责进行潜在的或内部的言语(即思想)",因此,华生曾经记录过人在思维时喉部和手部的肌肉活动,他坚信"肌肉习惯"就是"思维"。这一观点现在证实是有一定正确性的。现代思维研究逐渐揭示出思维往往伴随有躯体的运动,特别是与思维内容有关的肌肉的活动。例如一个人想象举手时,就可以从其手臂上记录到肌肉活动,尽管他的手臂不一定真的举起来。但是,思维中的躯体运动毕竟不是思维的主要方面,它主要还是受中枢控制的内隐过程。其实,华生也承认:思维的发展,也使言语的肌肉活动日益熟练、缩减,并过渡到内隐的过程。

行为主义对思维的研究,最典型的还是桑代克(Edward Lee Thorndike, 1874—1949)的动物解决问题的研究。桑代克为研究动物解决问题的过程专门设计了一个迷箱,被称为桑代克迷箱(Thorndike's pizzle box),如图 9-1 所示。实验时,将饥饿的动物放入箱内,观察动物如何打开箱子的门闩以逃出迷箱,获取箱外的食物。打开门闩的方式可以根据研究的需要设置成不同的难度。图 9-1中,绳子的一端固定在门闩上,另一端固定在箱内一踏板上。动物只要按压踏板,拉动绳子,抽出门闩,门就会被打开,动物就可以跑出箱子吃到食物。记下动物逃出箱子所用的时间和行为表现,再将动物放回箱内,让其进行下一次尝试。多次实验,直至最后动物能准确、快速地按压踏板逃出箱子。桑代克使用小猫等动物作被试,得到了动物随着尝试次数的增加,打开箱子所需时间逐渐减少的学习进程曲线,如图 9-2 所示。

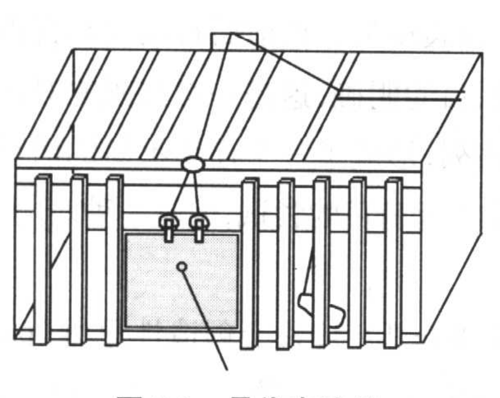


图 9-1　桑代克迷箱
(林崇德等:《心理学大辞典》,2003)

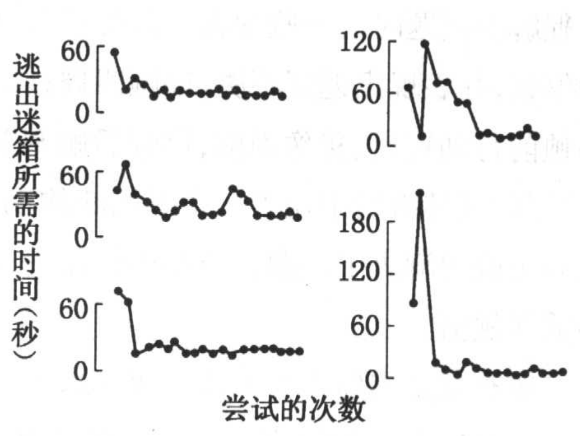


图 9-2　桑代克迷箱实验结果示意图
(林崇德等:《心理学大辞典》,2003)

桑代克根据大量的实验结果,认为动物解决问题的过程就是一个不断尝试和不断错误的过程。在尝试过程中,逐渐把问题情境、成功反应、行为结果强化联结起来,由此提出了学习的联结理论(connectionistic theory)。但是,他将动物实验的结果直接推广到人类问题解决情境是不太恰当的。

(二) 完形主义的研究

与行为主义同时代,始终站在行为主义对立面的完形主义对思维的研究更有成效,其影响也更深远。这主要体现在以下三个方面。

首先是关于思维过程的研究。按照格式塔心理学的观点,思维过程是:当个体环境里出现尚未得到解释的“紧张”时,就出现了“问题”,用我们现在的话来说,就是“思维发生于个体面临着一个不能以惯常手段解决的问题的情况下”。而思维的过程就是这种“紧张”解除的过程。解除紧张就要找到答案,而答案似乎是在对“紧张”的观察或知觉中。所以在完形主义心理学中,思维与知觉没有明显的界限,甚至认为思维就是知觉的一种形式,问题答案不是想出来而是看出来的。当然思维不是简单的知觉,它实际上是一种知觉重组,解决问题的人必须重组环境,即从几个不同的角度进行观察,直到事件之间的相互作用产生出答案的一个清楚图解。也就是说,人是在观察中看出事物各方面的整体联系,从而得到问题的解法或答案的。一旦完形出现,看出答案,紧张就解除。

其次是苛勒(Wolfgang Köhler, 1887—1967)关于问题解决的顿悟实验(insight experiment)。苛勒曾经受命在一个小岛上研究猩猩思维问题,后来第一次世界大战爆发,他在小岛进行研究的事几乎被政府部门遗忘,于是他在小岛上研究达 6 年之久。其中,对猩猩解决问题的研究最有影响。苛勒在观察中发现黑猩猩解决问题时的一些现象:(1)经过最初的尝试和失败后,常常出现很长时间的停顿,并表现出迟疑不决、环顾四周;(2)停顿表现为它们前后行动的转折点:停顿前行动盲目、犹豫困惑,停顿后顺利前进、目的明确,这是一个强烈的对比;(3)停顿或转折后出现了一个不间断的动作序列,形成了一个连续不断的活动,从而正确地解决了问题。苛勒由此提出,思维不是盲目的尝试,而是一种对情境的突然领悟。

还有就是韦特海默(Max Wertheimer, 1880—1943)关于创造性思维(productive thinking)的研究。韦特海默对创造性思维进行了系统研究,并于 1945 年出版了《创造性思维》一书。他研究的范围很广,从儿童解决简单问题一直到爱

因斯坦发现相对论的思维。他认为，创造性思维与对问题中某些格式塔的顿悟有关：打破旧的格式塔，发现新的格式塔，这就是创造性思维。韦特海默曾经对学生求平行四边形面积进行过考察。他发现，许多学生遇到稍有变化的平行四边形时，画不出符合要求的垂线。而要画出符合要求的垂线，就必须懂得垂线与图形面积之间的关系。这就说明，这些学生只知道要画垂线，而不知道这条垂线与整个问题情境的关系，不知道从整体上把握问题。

完形主义的整体观对后来的思维研究，乃至整个心理学的研究都影响深远。

四、信息加工心理学的研究

信息加工心理学(information-processing psychology)没有明确地指出思维是作为认知过程的某个阶段，但是这一理论在解释人的认知过程时，使用了编码、比较、匹配等概念，这里包括了对信息与信息之间联系的反映，就是思维的成分。信息加工心理学对思维的直接研究是使用计算机模拟(computer simulation)方法研究人的问题解决的思维过程，具体方法是：首先分析人类解决问题的过程，以此为基础编制计算机程序，再把这一程序在计算机上运行。如果程序能够运行并能得到与人类解决问题相似的结果，则证明这种模拟是成功的，这就说明对人类解决问题过程的理解是正确的；如果程序运行不能成功或不能得到与人类解决问题相似的结果，则说明未能真正理解人类解决问题的过程，于是对计算机程序进行修改，直至达到预期结果，再以成功实现的计算机程序来解释人的解决问题的思维过程。这种方法虽然具有逻辑倒置之嫌，但毕竟为思维心理学的研究开辟了一条新道路。1956 年，纽厄尔和西蒙提出了证明数理逻辑定理的程序，叫做“逻辑理论机”(Logic Theory Machine，简称 LTM)；1972 年，他们又创制了“通用问题解决者”(General Problem Solver，简称 GPS)。不过，直到目前为止，心理学家或人工智能专家也还没有真正地编制出一个能完全模拟人类思维的计算机程序。认知心理学用产生式系统(production system)描述人类思维和人工智能(artificial intelligence，简称 AI)。所谓产生式系统，是指计算机和人能执行的一组活动。只要有一定的条件就能产生一定的活动，其通用表达式就是“If . . . then . . .”。有一个条件，就可以执行一个活动得到一个结果，再有一个条件就再有另一个活动和另一个结果，依此类推。当条件输入完毕，活动执行完毕，就达到总的目标。如此看来，只要条件足够，问题都是可以得到解决的，但实

际上并非如此。比如,我知道某人在南京市的某家餐馆当大厨,那么我一个不落地把南京的餐馆都找遍,在理论上是可以保证能达到目的的,但在实际上,谁也不去这样做,或者就是这样做也不一定就真能找到这个人。所以,产生式问题解决理论并不能真正地解释人们的问题解决过程,或者说,不能完全解释人类的问题解决过程。

当然,信息加工心理学关于思维,特别是问题解决的研究,也不是像这里所表达的那样简单,它毕竟有了系统的理论和方法,并且推动了思维心理学的实验研究,我们将在后面作专门介绍,包括对两个划时代代表人物的介绍:纽厄尔和西蒙。

总的来说,思维研究还处在非常初期的阶段,其在心理机制和生理机制方面都还没有取得令人满意的结果,但是在科学研究、人力资源开发、教育训练、人工智能开发等诸多领域对思维研究的需要却越来越迫切。为了推动这一领域的研究,心理学家也尝试了一系列方法。下面,我们分别介绍概念形成与问题解决的实验研究及相关方法。

第二节　概念形成的实验研究

心理学中所讲的概念(concept)就是指一类事物的共有特征,这些特征可将其与其他事物区分开。因此,从这个意义上说,概念就是一种规则,即将事物进行分类的规则。概念形成(concept formation)就是发现这样的规则的过程。找到这一规则的最低表现形式就是能够正确地对事物进行归类,尽管有时不能用言语表达出来。比如用动物作为实验被试,虽然它们不能用言语表达自己对规则的反映,但是它们可以通过行为表现出它们对规则的把握。有研究者曾让猴子挑拣“长方条”来换东西吃。实验的方法是:在猴子的周围放着各种形状的木块,有方形的、圆形的、三角形的、长方形的,还有许多不规则的形状。当猴子拿来长方形的木条就给它一块糖果吃,凡是拿其他形状的木块都不给糖果。经过若干次实验,猴子学会了专门挑拣长方条来换糖吃,甚至在找不到长方条时拿着自己的尾巴来“冒充”,也不去拿其他形状的木条。这说明它发现了这一类木块共有的不变特征,这就是概念形成,也叫概念学习(concept learning)。关于概念形成的经典实验来自布鲁纳的研究。

一、概念形成实验的范式

（一）概念形成实验的材料

概念形成实验中，使用的刺激材料可分为两大类：人工概念（artificial concept）材料和自然概念（natural concept）材料。在实验室研究中，为了实验简单可行，并为了控制一些偶然因素，常常采用人工概念刺激材料。

赫尔（Clark Leonard Hull, 1884—1952）于1920年首次提出并使用了人工概念。他的人工概念是以汉字为材料的，即其概念样例是汉字，概念的定义是汉字的偏旁等特征，并以“oo”、“ta”、“yer”等字母组合作为概念名称，如图9-3所示。

图9-3显示，无意义音节“deg”对应的概念定义是所有包含偏旁“石”的汉字，其样例包括“砖”、“磐”、“研”等等；无意义音节“li”对应的概念定义是所有包含“力”的汉字，其样例包括“劬”、“動”、“劫”等等。不过，赫尔实验中，主试不是明确要求被试形成概念，而要求被试做的事情是机械识记，即学会将名称或无意义音节与呈现的样例联系起来。比如，给被试呈现一套卡片时，要求被试记住每张卡片上所标的是一个什么音节。如果被试不能注意到无意义音节总是与汉字的某一偏旁特征相联系，他就要花费更多的精力去识记；如果他能发现无意义音节总是与某一特征相联系，那么他就可以节省很多精力。被试是否能够发现概念，主要看他能否将某一音节运用到与其对应的一系列样例中。①

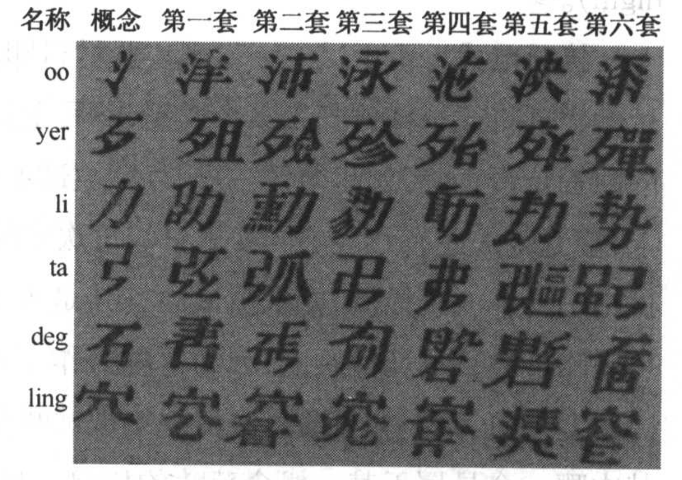


图9-3 赫尔实验使用的刺激材料

从程序看，赫尔的研究似乎不是真正的概念形成，因为这种概念形成具有偶然性，不是被试的“主要任务”。不过，赫尔的工作引发了一系列后续的研究，许多研究者按照各自不同的方法编制出各种各样的人工概念实验材料，为概念形成实验提供了极其丰富的材料，也使比较精确的实验室研究概念形成的过程成

① 邵志芳：《思维心理学》，华东师范大学出版社2001年版，第56—57页。

为思维心理学中一个独特的研究范式。

但是,20 世纪 70 年代以后,研究者对人工概念的形成实验产生了质疑,越来越重视自然概念形成的研究。所谓自然概念,就是自然世界中存在的概念。但是,在实验情境下,以自然概念为材料,同时保持个体概念形成过程的自然属性,就会消耗大量的时间,研究方式也就转换成了"观察法"或现场研究了。所以,到目前为止,实验室研究概念形成的方式还多是以人工概念为主。

(二) 概念形成实验的操作方式①

赫尔的实验操作方法并不为后来的研究者所采纳,因为概念形成在他的研究情境中实际上是作为识记加工的"副产品"加以研究的。后来的研究者采用较多的方式有两种:接受范式(reception paradigm)、选择范式(selection paradigm)。②

接受范式　由主试一一呈现实验材料即样例(sample),每呈现一个样例,被试就尝试将这一样例归入某个类别,然后主试给予被试正误反馈,即告知被试他的回答是否正确,如此循环往复,直至被试不再发生分类错误为止。在具体操作上,接受范式还有几种不同变式:(1)每次实验都将所有实验材料呈现给被试,主试指出其中一个样例,要求被试回答其是否属于某一概念范畴。这种实验程序可以降低被试的任务难度,因为这种操作下被试对该样例与其他样例的异同的比较变得更为容易。(2)每次呈现给被试两个或两个以上的样例,要求被试指出其中哪一个是属于某一概念范畴的样例。显然,这种方法也会降低被试的任务难度。(3)要求被试在主试给出反馈前或反馈后提出假设,这种方法可以收集被试实验中更丰富的口语材料,有利于对概念形成的思维过程进行更详尽的研究。

选择范式　由被试选择刺激材料作为某个概念的正例或反例,然后主试给予正误反馈,接着被试给出假设,如此循环进行,直到被试的假设正确、确定为止。

这两种方式实际上也反映了人们概念形成的实际过程,即在实际的概念形成中,既有选择性质的,也有接受性质的。当然,这两种过程可能存在内部机制上的差异,也存在与被试特征的交互作用,比如有研究发现,不同年龄个体在这两种实验范式中的表现有所不同:成人被试在选择范式中概念形成较快,儿童被试在接受范式中概念形成较快。

---

① 邵志芳:《思维心理学》,华东师范大学出版社 2001 年版,第 54—55 页。

② [美]John B. Best 著,黄希庭主译:《认知心理学》,中国轻工业出版社 2000 年版,第 335 页。

(三) 概念形成实验中的行为测量

根据概念形成实验的基本程序:呈现刺激(接受或选择)→被试反应(同时提出假设)→主试反馈("正确"或"错误")→(被试提出假设)→再呈现刺激。可以看出,实验中需要观察、测量和记录的行为主要包括:每一次刺激材料的内容与呈现方式、被试的反应及提出的假设、主试的反馈、被试在反馈后提出的假设、每一次使用的时间和尝试次数、整个实验序列的实验次数和所消耗的总时间。其中最基本的资料包括实验次数、需要时间、错误数等。

除上述行为测量外,主试还要在实验过程中注意观察被试的情绪状态及其变化、被试使用的主要策略等。

为了达到对被试上述行为的测量,并反映被试概念形成的过程及其影响因素,要在实验前编制并呈现指示语。概念形成实验的指示语一般要包含下列内容:(1)对概念作一般描述;(2)对刺激材料及其呈现方式的一般描述;(3)对被试反应的要求,要非常明确地告知被试如何反应、何时反应;(4)主试将给予被试什么样的反馈、反馈的意义是什么。

下面,我们以布鲁纳等人的经典实验说明概念形成实验的操作程序。

二、布鲁纳等人的概念形成实验

布鲁纳(Jerome S. Bruner, 1915—　),美国著名教育心理学家,当代认知心理学研究先驱之一。早年喜欢心理学,1937 年毕业于杜克大学,后入哈佛大学师从生理心理学家拉什利研究动物知觉。"二战"爆发后应征入伍,在艾森豪威尔总司令部下属的心理战部门工作,1945 年返回哈佛大学,1952 年任心理学教授。后来的生涯中主要从事教育改革领导与指导工作、教育心理学与认知心理学的研究等。

1956 年,布鲁纳与古德诺(Goodnow, J. J.)、奥斯丁(Austin, G. A.)共同完成的概念形成实验成为该领域的一项经典研究。

(一) 实验材料

布鲁纳等人使用的实验材料是如图 9-4 所示的 81 张卡片。这些卡片共有四个维度的特征,而每个维度又都分三个水平。

形状特征:圆形、十字形和方形;

颜色特征:红色、绿色和黑色;

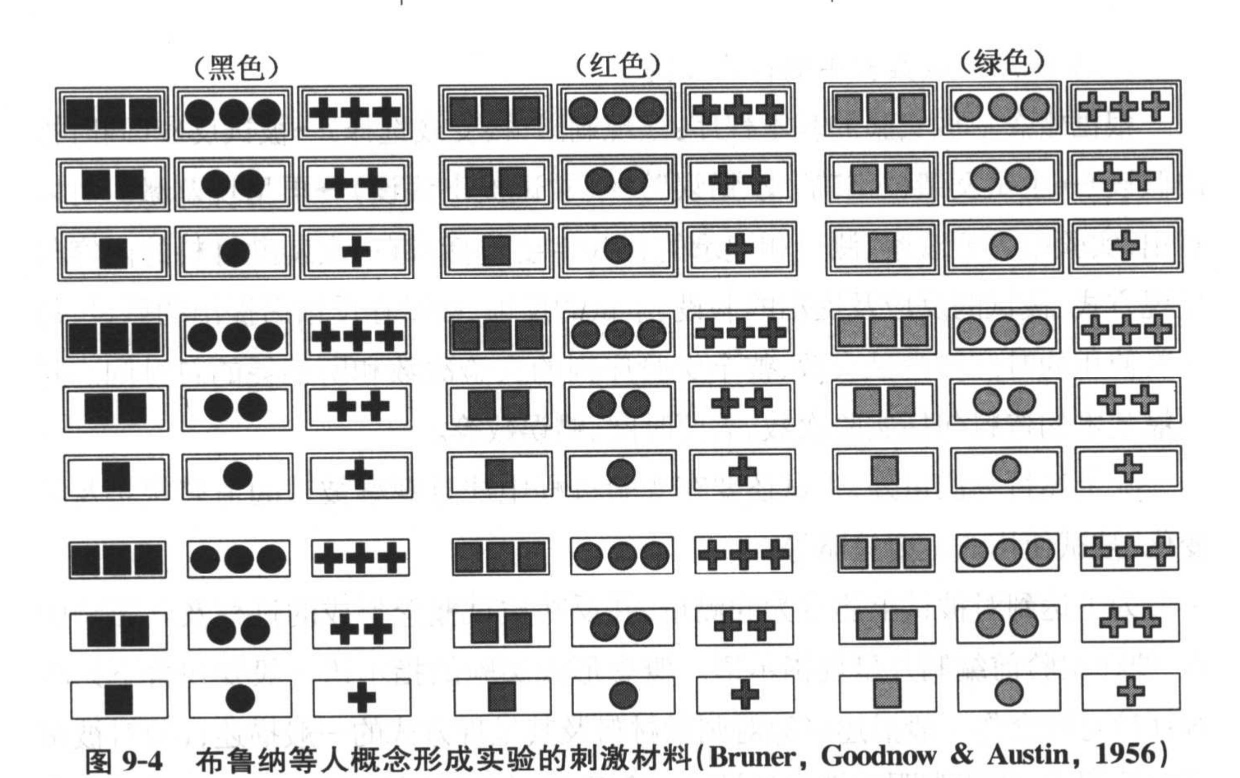


图 9-4　布鲁纳等人概念形成实验的刺激材料(Bruner, Goodnow & Austin, 1956)

边框特征：单线框、双线框和三线框；

图数特征：一个、两个和三个。

这一系列的卡片可看作是一个样例系列，其中包含大量的特定概念，每个概念都可以表述为一个规则。在这些卡片中，那些符合特定规则的卡片就可归入到一类，作为一个特定概念中的样例；不符合这个规则的，就应被排除在特定概念样例的范围之外。

(二) 实验程序

被试的任务就是确定某一卡片类别的分类规则。刺激呈现的方法既可以是接受范式，也可以是选择范式。比如以接受范式进行操作时，实验的操作程序是：

首先，研究者要根据研究的目的确定规则的类型。在研究中，研究者可以使用的规则类型主要包括：(1)单值概念(single-value concept)，要求被试确定具有某一特定维度的某一特定值的所有卡片都在其中的概念。例如，确定所有“双线框”卡片的规则就是一个单值概念。(2)合取概念(conjunctive concept)，规则确定了在一个维度上有一个取值，而在另一个维度上也有一个取值的概念。例如，印有“红色圆形”的卡片，这里确定了卡片在两个维度上的取值，一个是在颜色维度上取“红”，同时在形状维度上取“圆”，该概念规定了同时具有“红色”和“圆形”特征的所有卡片。(3)析取概念(disjunctive concept)，规则确定了

一个维度上的某一个值或另一个维度上的某个值,比如规则确定图形数量维度上的"2"值或边框维度上的"2"值,那么该概念就包含了图形为"2"或边框为"2"线的所有卡片。

确定了要求被试寻找的规则之后,即可向被试呈现刺激卡片,并要求被试作出适当的反应。这里既可以采取接受范式,也可以采取选择范式。采取接受范式时,主试向被试呈现一张一张的卡片,要求被试回答这张卡片是不是规则规定的概念样例;采取选择范式时,主试先给予被试一张卡片的刺激,然后让被试自己选择那些可能属于规则确定的概念的样例。在被试的每一次反应之后,要求其讲出规则,主试记录其使用的方法、尝试的次数和所用的时间等。直到被试正确地进行反应并能讲出主试规定的规则为止。

最后,再对实验记录的材料进行详细分析,以从中发现被试概念形成的特点和规律,特别是发现被试在学习概念时采取的策略、被试在概念学习时的情绪表现等。

(三) 实验发现

布鲁纳等人在上述概念形成的实验研究中,发现被试在接受范式及选择范式实验中,常常会使用一些不同的策略。

1. 接受范式实验中常用的策略

整体策略(wholist strategy)。使用这种策略的被试试图记住主试指出为"正确"反应的样例的所有共同属性或特征,而忽略掉其他不同的属性或特征。这也是一种排除法,即排除那些被主试肯定的样例所不具有的所有属性或特征。在接下来的实验中,接受新的卡片刺激和主试反馈,继续排除非共同的属性或特征,最后逐渐集中到概念规则所规定的卡片属性或特征上。

部分策略(partist strategy)。使用这种策略的被试一次实验中只预测并集中记住某一个假设或少数假设,如果能够正确预言这张卡片的类别就加以保持,否则就重新形成假设,直到获得一个总能被证实的假设,从中发现概念规则所规定的卡片的属性或特征。

布鲁纳的实验中,两种策略比较而言,整体策略较为理想,有65%的被试使用了整体策略。现在举例说明整体策略的使用方法:①

① [美]John B. Best著,黄希庭主译:《认知心理学》,中国轻工业出版社2000年版,第335—336页。

假如被试看到的第一张卡片是:"两条边框、两个红色的方形。"

一被试试图判定为合取规则,并假设他正确地猜测到这张卡片是概念,那么这张卡片就成为最初的肯定实例。现在被试试图编码所有的特征,即两条边框、两个图形、红色和方形。

接下来给被试呈现的卡片为:"一条边框、一个红色方形。"

被试将判断这张卡片不是概念。如果主试告诉被试他的回答是正确的,那么正确的假设将得到保持。如果告诉被试他回答不正确,这就是说第二张卡片是概念,那么被试就有可能在原有的假设和当前卡片的共同成分的基础上形成新的假设,即"红色方形",从而将两张卡片中的不同属性或特征排除,即将边框数、图形数特征排除。

当然,在其他研究情境中,如果刺激样例的属性或特征太多,整体策略未必最为有效。

2. 选择范式实验中常用的策略

在选择范式实验中,除使用上述整体策略和部分策略外,布鲁纳还发现被试使用下列四种有效策略。

同时性扫描(simultaneous scanning)。被试在实验前期构造出各种可能的假设,然后在每次实验反馈后去掉不合理的假设。

连续性扫描(successive scanning)。被试先从一个假设入手,如果他能依次正确地预测卡片,就保留假设,否则放弃该假设。

保守性聚焦(conservative focusing)。被试在实验前期,先进行猜测直至正确击中一个假设为概念的卡片,然后选择其他卡片,选择中每次只变换最初的肯定样例中的一个特征。

博弈性聚焦(focus gambling)。被试在实验前期,先进行猜测直至击中一个假设为概念的卡片,然后选择其他卡片,但在选择中每次都变换最初的肯定样例的两个以上的特征。这种策略运用得当时,可以缩短探索的时间,但也有可能会因失败而出现思维混乱。

概念形成实验中,被试会在具体样例的反应与反馈中,灵活采用多种不同的策略或同时采用多种策略。

在系列实验研究的基础上,布鲁纳提出了概念形成的"假设检验说"(hypothesis-testing theory)。该假说认为,概念形成是一个不断提出假设又不断检

验假设的过程。在这个过程中,被试根据自己的知识经验对刺激材料进行分析与综合,提出概念假设,然后再根据主试的反馈和对新材料的分析,检验和修正假设,如此反复进行,直至最终形成概念。20 世纪 60 年代,莱斯特尔(Restle, F.)根据布鲁纳的理论,提出了概念形成的"假设检验模型"(hypothesis-testing model),如图 9-5 所示。[①]在这个模型中,包含以下两个主要观点。

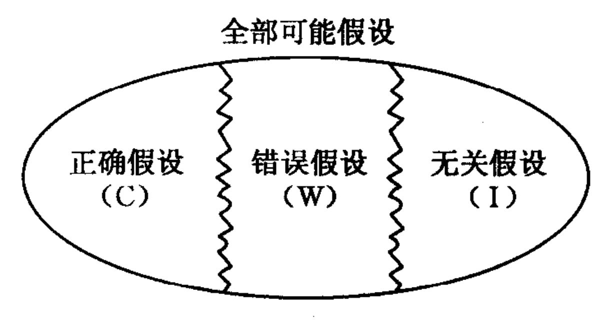


图 9-5 概念形成的假设检验模型
(汪安圣,1992)

(1) 抽样观点:被试从所有可能的假设全集中抽取一个(或一组)假设,它可能是正确的(C)、错误的(W),或无关的(I)。

(2) 无记忆观点:如果选出的假设正确,被试就可以作出正确的反应并保持这个假设;如果假设是错误的,就会作出错误的反应,假设就会被放回到原处,下一次还有可能被抽取到;如果假设是无关的,它既可以导致正确的反应,也可以导致错误的反应,因此既可能被保持,也有可能被放置到原处并还有可能被抽取到。比如,"红色"是概念的特征,那么"红色"的假设就是正确的,"绿色"的假设是错误的,"方形"的假设就是无关的。

这一概念形成模型得到后续许多研究的支持,确实能在相当程度上反映人们概念形成的心理机制,是迄今最具有解释效度的概念形成理论模型。

三、"耶克斯选择器"与空间位置关系

在心理学实验室里,有一个常规的实验设备,叫做"耶克斯选择器",它可以用来研究空间位置关系概念的形成过程与规律。

(一) 耶克斯选择器

耶克斯(Robert Mearns Yerkes, 1876—1956),是美国比较心理学家、实验动物心理学研究先驱,在心理测验方面具有丰富的理论与实践经验,编有"概念形成的复合选择测验",后人以此为基础设计出测验空间位置关系概念形成过程

① 汪安圣:《思维心理学》,华东师范大学出版社 1992 年版,第 89—90 页。

的仪器,就叫做"耶克斯选择器"(Yerkes selection box),①如图 9-6 所示。该仪器有 12 个活动电键,声音可按主试的意图与任一电键相连;中间有单向屏风,不让被试看到主试的操作,主试却能看到被试的反应。

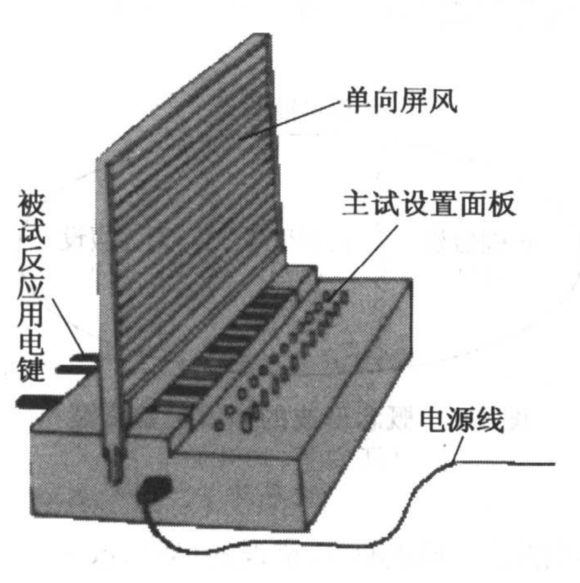


图 9-6 耶克斯选择器
(杨博民,1989)

(二)实验任务与程序

以简单位置关系概念形成的实验为例。实验中,主试为被试设定一个空间位置关系刺激模式,即将某些电键推出去构成一个空间位置刺激模式,同时设定这推出的若干电键中有一个或两个电键是与声音信号刺激联系在一起的,而且这一个或两个与声音联系的电键在空间位置上符合某一确定的规则。被试的任务就是在主试不断地改变刺激模式的时候,找到其中隐含的不变规则,这是被试在实验中的基本任务。

实验的一般程序是:

第一步,主试先准备好让被试探索的空间位置关系概念,即声音和哪一个符合一定空间位置关系的电键相联系,并按这一规则设计出一系列的刺激变式,排好呈现顺序。

第二步,让被试坐在单向屏风前,主试在屏风后操作:按照事先设计好的方案推出几个活动电键,并使其中一个与声音相通(不要让被试知道),然后向被试交代任务,即呈现实验指示语:"这些电键中有一个是与声音相联系的,请你找出是哪一个。找到后请你记住声音是与什么位置的电键相联系的。第二遍我换几个电键,其中也有一个与声音相联系,你的做法与上一次相同,如果你能知道哪一个是与声音相联系的,就直接去按这个电键,如果还不确定,你就需要探索。如此进行下去,直到你能连续三遍第一下就按对了,并且能正确地说出声音键的空间位置关系为止。你的任务就是较快找出每一次实验中与声音相联系的那个键与其他键的位置关系。虽然每一遍给你的电键数量及位置均不同,但要你找

① 国内较为理想的"耶克斯选择器"为北京大学教学仪器厂生产。

的空间位置关系是相同的。在你每次按键的前后说出你自己的想法。”①

主试记录被试实验进行了多少遍、每一遍所用的时间(从主试喊“开始”,同时按动秒表计时。当被试按键发出声音时,主试按动秒表停止计时)、实验中的口语和表情等。

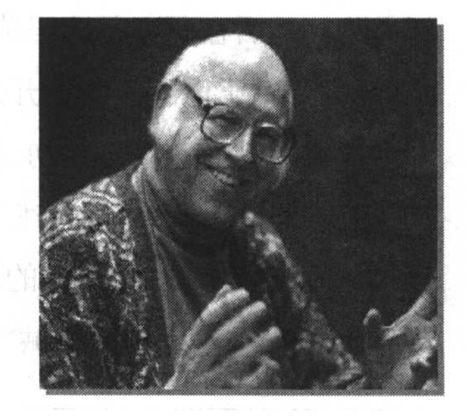
第三步,更换被试进行实验,直到所有被试都完成实验。对记录的实验资料进行定性和定量相结合的分析,概括出被试概念形成的过程、规律及被试使用的策略等。

第三节　问题解决的实验研究

如前文所述,符兹堡学派、行为主义学派和完形主义学派先后都曾就问题解决进行了专门的实验研究,并提出过各自的问题解决理论。但是,现在看来,这些研究都还比较表面化,未能深入其深层心理机制或对深层机制的解释较为含糊。20 世纪中期以后,该领域出现了新理论和新方法——信息加工的问题解决理论与方法,随后,问题解决研究成为思维心理学范围内最活跃的专题。

一、纽厄尔和西蒙

不管是以“问题解决”为关键词,还是以“人工智能”为关键词进行文献搜索,都能高频率地看到一个“三人组合”,那就是著名的认知心理学家、人工智能专家艾伦·纽厄尔(Allen Newell, 1927—1992)、赫伯特·西蒙(Herbert A. Simon, 1916—2001)和克利夫·肖(Cliff Shaw),在他们的周围还聚集了一批勤奋的同事和青年学生。纽厄尔和西蒙都是公认的划时代的开拓者。



艾伦·纽厄尔

(一) 纽厄尔的研究生涯与主要贡献

纽厄尔于 1927 年 3 月 19 日出生在美国的旧金山,其父亲罗伯特·纽厄尔是斯坦福医学院的杰出放射学教授,对纽厄尔的

① 杨博民:《心理实验纲要》,北京大学出版社 1989 年版,第 320 页。

事业发展提供了榜样作用和重要影响。纽厄尔中学毕业时，正值"二战"结束，他开始在一个造船厂里工作，随后被美国海军征召入伍。在海军服役期满后，他进入斯坦福大学读书，攻读物理学专业，大学毕业时完成了他的第一篇研究论文，是关于 X 射线光学的。在斯坦福，纽厄尔结识了当时著名的数学家乔治·波利亚(George Pólya, 1887—1985)。波利亚的一本著作《如何解决》(*How to Solve It*)(1945)向他介绍了"启发式"(heuristic)这一"发现的艺术"(the art of discovery)。纽厄尔从中知道了"启发式"或发现的艺术是可以被研究和分析的，而且知道它在创造性思维中起到非常重要的作用。

1949 年至 1950 年，纽厄尔作为研究生在新泽西州的普林斯顿大学城从事数学研究，主要致力于游戏理论研究。这一研究使他确信可以在纯数学的研究中把实验方法和理论方法结合起来。离开普林斯顿后，他进入 RAND 公司①并被编入空军后勤研究小组，开始进行组织理论、决策等方面的实验研究。1952 年与同事一起建立"系统研究实验室"(Systems Research Laboratory)，随后派生出"系统发展公司"(Systems Development Corporation)。公司研究的中心问题是记录和分析雷达监控小组成员与雷达屏幕、雷达侦察机及其他因素之间的交互作用。有了这些记录及分析资料，纽厄尔将注意力集中到雷达站工作人员处理信息和进行决策的过程方面，并进而寻找分析和建立过程模型的适当技术。这时西蒙作为实验室的顾问与纽厄尔相遇。在第一次相遇中，他们彼此找到了研究的共同兴趣、背景或基础：信息加工研究可以为理解人类组织决策开辟一条新道路。随后在研究雷达屏幕上模拟飞行物的图像时遇到了程序设计员肖。自此，他们开始了以人工智能开发为主的合作研究。

早在 20 世纪 50 年代初，神经机械学和人工生命的思想已有所传播。1950 年，图林(Turing, A.)等描述了用计算机下棋的想法；1952 年，罗斯·阿什比(Ross Ashby)出版专著《脑的设计》(*Design for a Brain*)；1953 年，格雷·沃尔特(Grey Walter)发明的机械"龟"可以在房间中进行四处搜索并找到房间出口；

---

① RAND 公司(RAND Corporation)最早成立于 1946 年，当时是作为道格拉斯航空公司(the Gouglas Aircraft Company)的智囊机构，RAND 是 Research And Development 的缩写，是第一个被称为"智囊团"(think tank)的机构，也是第二次世界大战的产物。1948 年，RAND 公司从道格拉斯公司中分离出来成为一个独立的、不以盈利为目的的咨询公司，其工作的宗旨是：RAND is a nonprofit institution that helps improve policy and decision making through research and analysis. 详情可到下边网站查阅：*http://cognet.mit.edu/MITECS/Entry/Simon2*.

1954年,纽厄尔和西蒙在一次旅途中长时间地讨论了用计算机模拟人类问题解决的可能性。我们可以把1954年看作是计算机模拟人的问题解决思想萌芽的时间,也就是人工智能开始酝酿的时间。由此,信息加工问题解决理论开始逐步形成。因为正是人工模拟这样一种非常明确的思路或目的,才使得问题解决研究的理论与方法尽可能向着迎合可模拟性的方向发展。

就纽厄尔的研究转向来说,1954年确实是非常重要的一年。这一年他在RAND公司参加了一个研讨班,聆听了来自林肯实验室的奥利弗·瑟尔福里奇(Oliver Selfridge)教授的报告,即关于一个能学会识别字母和其他图形的计算机程序的报告,这是一个相当初级但却可运行的程序。这次研讨班的经历对纽厄尔的影响很大,使他强烈地意识到:"智力适应系统是可以建立的,尽管它可能比已经做过的任何事情都要复杂得多。"①于是,他开始将全部精力投入到通过模拟来理解人类学习和思维的研究中去。几个月后,纽厄尔撰写的文章《象棋机:一个借助自适应来处理复杂任务的范例》②描述了富于想象的类似于人下棋的计算机程序设计思想。虽然他自己没有将这一思想付诸实施,但却被1958年的NSS③象棋程序所借用。而此时纽厄尔的目标已经超出了象棋程序的开发:"因此,这种努力的目标是制成一种通用的能够下'一手'好棋的计算机,这是更好地理解能够处理极端复杂问题的各种计算机、机制和程序的一种手段。"④可见,这时纽厄尔已经具有了比较明确地采用人工模拟的方法来理解问题解决内在机制的思想,可以认为这是人工智能信息加工计算机模拟的早期思想或研究取向,也是物理符号系统开始形成的标志。

1955年,纽厄尔来到匹兹堡与西蒙和肖一起工作,并在匹兹堡获得工业管理方向的博士学位。由于各种技术和偶然的原因,他们研究的兴趣从象棋转移到几何学,又从几何学转移到逻辑和逻辑理论机(Logic Theory Machine,简称LTM)。为使前述的设计思想得到实现,他们同时在寻求一种合适的程序语言,这就导致了信息加工语言(Information Processing Languages,简称IPLs)的提出,这是最早的

---

① 摘自 Herbert Simon 为纪念 Allen Newell 发表的演讲。网站:*http://cognet.mit.edu/MITECS/Entry/Simon2*.

② Newell, A. (1955). The chess machine: an example of dealing with a complex task by adaptation. *Proceedings of the Western Joint Computer Conference*.

③ NSS是指Newell、Simon和Shaw组成的"三人小组"。

④ Newell, A. (1955). The chess machine: an example of dealing with a complex task by adaptation. *Proceedings of the Western Joint Computer Conference*.

为计算机设计的表列加工语言(list-processing language)。“完全可以说,逻辑理论机(LTM)及其后继者通用问题解决者(General Problem Solver,简称 GPS),奠定了其后 10 年中大部分各种人工智能程序的基础。”①在纽厄尔看来,LTM 不是一种“演绎机器”(deduction machine),而是从假设的规则到一般原理进行反向的、归纳的推理,它发现规则就像发现其他东西一样是一种有选择的搜索。

LTM 及其“后继者”GPS 的研究和开发意义深远,因为它们并不是简单地针对某一单一任务的,它们都是属于通用的计算机程序,是能够解决一类逻辑问题的,而且“具体程序都是通向具有更普遍意义的一个问题解决的方法的一个步骤:理解人类的意识”。②

(二) 西蒙在问题解决领域的卓越贡献

毫无疑问,西蒙在当代认知心理学和计算机领域都是让人倍感亲切的人物,而他在经济学领域又是诺贝尔奖获得者,这真是让人感到不可思议的事情!

赫伯特·西蒙

西蒙出生于 1916 年,1943 年获政治学博士学位,1978 年因在经济学领域的突出贡献获诺贝尔奖。20 世纪 50 年代,成为计算机科学、人工智能研究和心理过程的计算机模拟等方面的开创者,也因此在 1974 年与纽厄尔一起获得美国计算机协会颁发的“图林奖”(Turing Prize)。②他在青少年时代是一个酷爱运动、勤于思考的少年,而他的嗜书如命更是让人难以置信。在步入科学研究的历程中,他表现出卓越的领袖才能,具有非凡的亲和力,因此在他周围聚集了一批来自多学科的研究人员和青年学生,特别是他与纽厄尔和肖的合作产出了改变当代认知心理学的划时代成果。

在西蒙的研究历程中,其前期主要集中于管理科学、社会学和经济学方面。但从 50 年代后半期开始,随着计算机技术的迅速发展,以及与纽厄尔、肖等人的合作,思维、问题解决和计算机模拟等构成他研究的主要内容。如果把他关于其

①② 摘自 Herbert Simon 为纪念 Allen Newell 发表的演讲。网站:*http://cognet.mit.edu/MITECS/Entry/Simon2*.

② Turing, A. M., 1950 年在《心灵》杂志上发表论文《会计算的机器和智慧》,界定了人工智能的领域,提出了认知科学的纲领,成为现代人工智能研究最重要的开创性人物。详见:Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX, 236.

他认知问题的研究也算上,他在认知心理学方面的文献很丰厚,内容也涵盖了认知心理学的几乎所有方面。所以,西蒙为现代信息加工认知心理学的建立和发展作出了杰出贡献。

纽厄尔和西蒙在问题解决理论方面的建设性工作主要是在20世纪50到60年代的十几年时间中。在这期间,他们不仅为人类问题解决研究积累了大量实证资料,也为该领域理论的建立创立了独特的语言和表达方式,一般可将其在问题解决方面的理论称为产生式问题解决理论。这种产生式理论与人工智能程序语言在本质上是一致的。关于西蒙的具体研究,此处不赘述,我们在介绍了问题及问题解决的基本概念之后,专门介绍西蒙等人建立起来的信息加工问题解决理论,以及他们在这一研究中采用的主要方法。

二、问题及问题解决的概念

(一) 什么叫做问题

当人们不能采取显而易见的直接手段达到目的时,就有了问题,它其实就是人们面临的一种认知情境,在这种情境中人们要达到目标存在某些障碍,所以问题的存在需要以下三种基本成分。①

成分一:给定(givens),是一组关于问题条件与问题情境的描述,即问题的起始状态;

成分二:目标(goals),是关于问题结论的描述,即问题所要求的答案,也就是问题的目标状态;

成分三:障碍(obstacles),从问题的起始状态到目标状态之间的某些未知的中介状态与各个步骤。因为这些中介状态和连接这些状态之间的步骤未知,所以就形成了从起始状态到目标状态间的障碍。

"总之,问题是给定的信息和目标之间有某些障碍需要加以克服的情境。在达到目标之前,可能会有一些错误和曲折,有许多中介的步骤。"②

(二) 问题的类型

格里诺(Greeno, 1978)把问题划分为以下三种典型的形式。

① 梁宁建:《当代认知心理学》,上海教育出版社2003年版,第276—278页。

② 汪安圣:《思维心理学》,华东师范大学出版社1992年版,第99页。

(1) 归纳结构问题(problem of inducing structure):给出几个条件,而问题解决者必须发现隐含在条件中的结构模式才能解决问题。例如,以下的数列填空问题、类推问题等。

问题举例 1:2、8、18、?、?、72,问号处应该填什么数字;

问题举例 2:镰刀——茅草、渔网——鲇鱼、钢笔——?、卡车——? 问号处应该填什么?

(2) 排列问题(problem of arrangement):给出某一事物构成的各种成分,问题解决者必须以某种合理的方式对这些成分进行排列,才能达到问题的目标状态。

```
    D O N A L D
+   G E R A L D      D←—5
———————————————
    R O B E R T
```

图 9-7 纽厄尔密码算术题
(Newell & Simon, 1972)

问题举例 3:纽厄尔密码算术问题。如图 9-7 所示,已知加法算式中的字母 A、B、D、E、G、L、N、O、R、T 各自代表 0~9 中的一个不同的阿拉伯数字。已知条件是 D=5,那么其他各个字母代表的数字是多少时,这个加法算式才是成立的?在这个问题中,字母代表的数字的全体已经给出,只是其如何在这个加法算式中排列是未知的,所以叫排列问题。

(3) 转换问题(problem of transformation):给出问题的起始状态和目标状态,而且问题结构成分均已知,只是其存在的状态不一样。最典型的例证是"传教士与野人过河问题"、"海诺塔问题"、"八瓦游戏问题"。

问题举例 4:传教士与野人过河问题。一天,三个传教士和三个野人同时来到渡口准备过河。渡口只停泊着一条小船,这条小船最多一次只能承载两个人,而且因为没有船工,所以过河的人要自己划船。因为野人比较野蛮,如果在河两岸的任一边出现野人数量超出传教士数量的话,传教士就有被野人吃掉的危险。请问,这六个人该如何渡过河去呢?

问题举例 5:海诺塔问题。如图 9-8 所示,在一木制底座上有三个立柱 A、B、C。在立柱 A 上套着五个圆片。要求用最少的搬动次数把这五个圆片搬到立柱 C 上去,在搬动过程中,限制每次只能搬一个、可把立柱 B 作为中转,而且只允许小的圆片压在大的圆片上,不能以大的圆片压在小的圆片上。那么如何来搬动圆片呢?

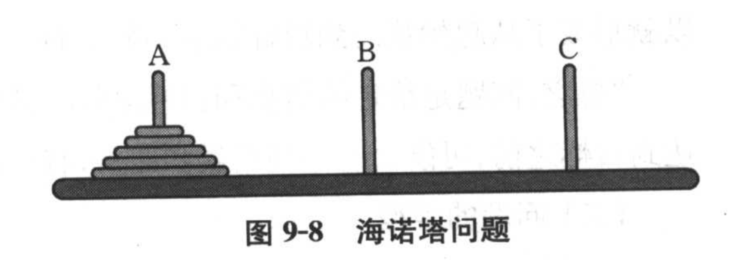


图 9-8 海诺塔问题

问题举例6:八瓦游戏问题。如图9-9,将1～8个数字写在八块瓦上,按图中左边的顺序摆放构成问题起始状态。要求问题解决者在3×3的方格内移动瓦片,每次只移一个,而且不允许移出方格。问最少要移多少次才能将八瓦摆放顺序转变成图中右边的样子?

起始状态

2	1	6
4		8
7	5	3

→

目标状态

1	2	3
8		4
7	6	5

图9-9 八瓦游戏问题

瑞特曼(Reitman, 1965)则根据问题条件与目标状态的界定清晰性把问题划分为规定良好的问题(well-defined problem)和规定不良的问题(ill-defined problem)。还有的心理学家根据问题解决是否依赖专门知识,把问题划分为知识贫乏领域问题(knowledge-lean problem)和知识丰富领域问题(knowledge-rich problem),或者分别叫做人工问题、专门领域问题等等。

(三) 什么叫做问题解决

安德森(Anderson, 1980)对问题解决的定义在认知心理学中具有代表性。他认为问题解决具有以下三个特征:具有明显的目的性,是一个操作序列,包含认知性的操作序列。因此,我们将问题解决的含义概括为:问题解决就是有目的的认知操作序列。

在认知操作过程中,随着问题解决者采取某种措施,问题的状态就会发生不断变化。把这些状态按照操作的顺序排列出来就构成一个由各种相互连接的问题状态分布图,心理学中把这个图叫做问题行为图(problem behavior graph)或问题空间(problem space)。所谓问题空间,就是问题解决的中介状态以及实现这些状态转换的操作方法。实现问题状态转换的操作方法叫做算子(operator)。问题空间是由算子连接问题中介状态构成的。

三、信息加工的问题解决理论

纽厄尔和西蒙的问题解决理论及其表述开始于20世纪50年代中期,成型于70年代初,被完整表述在他们的长篇著作《人类问题解决》(*Human Problem Solving*)中。①这本专著可被称为现代认知心理学关于问题解决研究的经典之

① Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N. J.: Prentice-Hall, 787—868.

作。西蒙在此书中直截了当地说："我们关于人类思维与问题解决的理论就是把人看作一个信息加工系统(IPS)。"[①]而构成西蒙问题解决理论体系的要素主要包括三个：问题解决者(problem solver)、任务环境(task environment)与问题空间(problem space)。[②]

(一) 问题解决者

在纽厄尔与西蒙看来，问题解决者作为一个完善的物理符号加工系统，除具有完善的感受器(receptor)与效应器(effector)外，他的主要结构包括记忆系统(memory system)和处理器(processor)，物理符号或说符号系统是其加工的对象。符号就是一系列具有象征意义的替代物，这些替代物表征实例或某事物的等价物等。当这些符号按照各种不同的关系联结起来构成一整体时，就叫做符号系统或符号结构系统，符号结构也因此可以表征被指代物的关系。看来，符号就是信息的载体，是人们可将其相互区分开来的各种模式，如文字、图像、声频、味觉等。有了符号的概念，人脑与电脑就有了类比的基础，它们都是物理符号系统！那么，究竟什么样的系统才算作物理符号系统呢？

物理符号是信息的载体，而进行信息加工则是智能的本质，因此人脑和计算机都是具有智能的结构系统，都被称为完善的物理符号系统，二者都具有下述六个功能或特征。

第一，输入(input)：如把字母敲进计算机、用眼睛看字等；

第二，输出(output)：如显示器显示的内容、用手写出字、用嘴说出话等；

第三，存储(store)：如用媒介记录信息、人记住事件或概念等；

第四，复制(copy)：如计算机复制文字、人认出某对象并复制成一表象等；

第五，结构化(build symbol structure)：如计算机找到符号间的关系以建立符号结构。用计算机自动作出一个统计图表来就是其一个结构化过程；人可将各种材料按照某种关系组织成一个系统，如心理现象的分类系统等；

第六，条件性迁移(conditional transfer)：依靠已有的符号可以完成后续活动，如计算机可以在一定符号储存的基础上完成一系列操作、人在已有知识经验基础上和新的刺激作用下可以发动和完成一系列动作。

① Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N. J.: Prentice-Hall, 19—20.

② 朱新民、李亦菲：《架设人与计算机的桥梁》，湖北教育出版社 2000 年版，第 92—94 页。

对于上述“符号”概念和“物理符号系统”的理解需要注意三点:(1)“符号”是一类物理事件的抽象形式,而不是实在物,因此它可以各种不同形式来表现。如“红色”可以由一朵花表现、可以由电子计算机的一组电子管来表现、可以由人的一系列神经活动来表现;(2)上述六个特征是完善物理符号系统的充要条件,具有这些特征的都是完善的物理符号系统,凡完善的物理符号系统都具有这些特征;(3)人脑具有智能,能进行信息加工,是完善的物理符号系统;计算机是完善的物理符号系统,它具有智能,这是人工智能的前提;人脑和计算机都是完善的物理符号系统,所以可以实现人的智能的计算机模拟。人脑和计算机具有功能上的相似,但它们可能以完全不同的形式来完成相同的功能。如果可以按照人类的思维过程来编制计算机程序,那么我们就可以使用计算机的信息加工形式描述人的认知过程或建立一个理论来说明人的认知。这就是信息加工心理学的基本意图和思维逻辑!

但是,人作为一个信息加工系统,其资源是有限的,即人作为信息加工系统具有一系列有限性,这些有限性制约了其信息加工的资源,这也是我们理解人类问题解决心理机制必须要考虑到的因素。概括地说,人的有限性主要包括:(1)注意的分配机制,如人的大脑必须对刺激信息进行过滤,在同一时刻只能重点注意一个事物,而这一事物往往是重要或新奇的;(2)记忆系统的限制性,如短时记忆和工作记忆的容量有限、信息的存贮和提取速度有限;(3)记忆信息的组织方式,知识组织具有组块化或“图式”(schema)化特征,这虽然降低了信息加工对资源的要求,但它依然是有限的,而且在有些问题解决者那里,这种有限性更明显;(4)思维中的搜索机制,人的思维往往是在很大的可能性空间中进行搜索,而这种搜索具有很大的选择性,很多时候也缺乏严密性,用于搜索问题空间的算子有时还是错误的;(5)序列加工与平行加工,人同时具有序列加工(serial processing)和平行加工(parallel processing)的能力,西蒙并且认为,在低级感知输入或运动支配过程中有许多平行加工,但在大脑皮层水平的记忆、思维、注意等过程则多是序列加工的。

有了这些限制性,与计算机相比,人在信息加工过程中,既有优越于计算机系统的主动性、灵活性,也有在信息加工资源方面的有限性,所有这一切决定了人作为问题解决者与问题情境之间的交互作用方式。

(二) 任务环境

任务环境是指某个特定问题的一切可能的知识状态的集合。这种知识状态是通过将有效算子(或操作)作用于每个问题状态而生成的。

任务环境包括问题呈现的初始状态、要求达到的目标状态、中间可能遇到的中介状态以及可以使用的算子和对算子的约束等。对于设计者或主试来说,任务环境“完全呈现在眼前”,因为他们对问题的各种状态与状态之间的转换算子非常清楚、“一目了然”。但对于问题解决者或被试来说,任务环境只是“部分呈现在眼前”,在其成功解决问题之前,任务环境处于半隐藏状态,需要问题解决者依靠现有线索与内部知识图式之间的相互作用,去“发现”任务环境的其余部分。

很明显,任务环境既然包含了所有可能的中介状态,也就包含正确的解决路线,但关键的是,这一路线对于问题解决者来说是未知的,被试的问题空间与任务环境可能一致,也可能不一致。

(三) 问题空间

问题解决者从刺激情境中提取信息后在头脑中建立相应的问题状态表征并操作算子实现问题状态的转换,如前文所述,这些问题状态和采用的算子构成了关于问题的心理结构,这一结构叫做问题空间。被试一旦建立起问题空间,他就会依据问题空间来进行解题操作,在问题空间中进行中介状态与算子的搜索。因此,初期形成的问题空间对问题解决行为有决定性的影响,甚至决定了问题解决的成败。

概括地说,问题解决是问题解决者与任务环境相互作用的过程。在这一过程中,问题解决者首先将任务环境表征为记忆中的问题空间,然后在问题空间中执行算子,连接问题的起始与目标状态,解决问题。当然,如果问题解决所建构的问题空间与任务环境不一致,也就是说,问题解决对问题的心理表征是错误的,那么在这样的问题空间中进行搜索是找不到问题正确解的,因为搜索的范围并不包括正确的解题路线。至此,我们知道了:问题空间决定着问题解决行为,以及这一行为是否会导致问题得到解决。但是问题空间是内隐的,所以需要借助特殊技术发现问题空间,才能有效地揭示问题解决的心理机制,这种技术就是口语报告分析法!

四、口语报告分析与问题行为图

“口语报告分析”之英文“protocol analysis”中的 protocol 取“与某附属物、某种说明或记录材料有关”之意,因此“protocol analysis”可直译为“关于描述直接经验或知觉的句子的分析”[1],在思维心理学领域,它与“出声思考”(thinking

① 廖伯琴:《中学生物理问题解决的表征差异及其成因探析》,西南师范大学博士学位论文,1999:19。

aloumd)同义。因为"口语报告分析"已成为思维心理学中迄今最有效的研究方法,在问题解决研究中被广泛采用。这种方法在心理学研究中的应用始于早期实验心理学,即从冯特的"自我观察法"开始,随后符兹堡学派套用冯特的方法来研究思维。不过,从这些早期的"自我观察"到埃里克森和西蒙的《口语报告分析法》(Ericson & Simon, 1984),它已经在具体的操作和使用中发生了许多变化。近年来,国内问题解决研究中又有多种具体的操作手法出现(李亦菲等,1998;廖伯琴,1999;刘晓晴等,2000;邓铸,2002, 2004)。毫无疑问,口语报告分析虽然是问题解决中最为有效的方法,但它又具有相当的操作难度,这难度不仅表现在可靠口语材料的获取过程中,也表现在对已经获取的口语材料的分析过程中。

(一) 早期的"口语报告分析"法

口语分析是从哲学内省法(introspection)生长出来的,"内省法是我们必须首先并且总是相信的观察方法,它使我们能窥视自己的心灵,并报告在心灵所发现的东西"。[①] 冯特则否定了这种哲学范畴中的内省法,将其嫁接在自然科学的实验控制过程中,形成了"自我观察法"(method of self-observation)或叫做"实验内省法"(method of experimental introspection)。铁钦纳发展了冯特的实验内省法,将被试训练加进实验过程。铁钦纳主张对实验被试进行严格的训练,使之掌握感知和描述自己意识状态的方法,以便能在实验中"仔细地分析心理的结构,把心理过程的基本元素从意识的纠缠中区分出来。"[②]

符兹堡学派在思维研究中独树一帜,他们使用的研究方法也包括口语报告。以屈尔佩为首的符兹堡学派在实验中"给被试一个词和关于这个词的一个问题,要求被试对这个词进行自由联想,然后要求被试描述导致反应的思维过程,或者要求被试描述回答某些问题的思维过程"。[③] 显然,符兹堡学派使用的是回溯性的口语报告分析法。但是由于符兹堡学派实验中使用的问题都属于简单问题,在许多时候,被试的思维过程难以展开,所以实验中得到的口语报告材料相当贫乏,实验研究因此未得到太多有影响的结果。

行为主义代表人物华生反对内省,但是他又认为人具有一种观察自己身体内部所发生的变化并对这些变化进行口语报告的能力。因此,华生也把口语报

① James, W. (1890). *The principles of psychology*. New York: Holt.

② 林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社 2003 年版,第 1132—1133 页。

③ 汪安圣:《思维心理学》,华东师范大学出版社 1992 年版,第 20 页。

告作为心理学研究的客观方法之一,他认为通过让被试对特定问题的出声思维,比起依靠不科学的方法来,能够更多地了解和认识思维心理学。不过,华生关注的不是被试报告其体验和意识,而是报告其内部的生理变化。武德沃斯(Robert Woodworth, 1869—1962)虽然没有明确表达是赞成还是反对使用口语报告,但他感到非常困惑的是,为什么"十秒钟的思维有时需要十分钟才能说清楚"。

1945 年前后,录音机开始投入使用,它可以实现口语报告的适时记录。1945 年,德国心理学家邓克尔(Duncker, K.)首先利用口语报告分析方法研究了人的问题解决过程,这大概是最早明确使用口语记录方法研究问题解决过程的心理学家。他根据被试的口语报告,将人的问题解决过程分为一般范围、功能性解决和特定性解决三阶段。在一般范围阶段,被试重述问题,提出大致的解题途径;在功能性解决阶段,解题范围缩小,被试对保留下来的解题途径作深入探讨和更具体的思考,提出关于问题的解决方案;在特定性解决阶段,被试要进一步探索某一功能性解决的具体方法。很明显,邓克尔根据被试的口语报告将问题解决过程划分成的三个阶段大致与我们今天所说的解题过程一致:提取问题的情境信息以表征问题,形成解题的基本思路——进一步分析与明确问题,将解题的操作范围缩小,提出明确的解题方向——将解题方法具体化,完成解题过程得到答案。

那么,口语报告分析法是如何实施的呢?

(二)"口语报告分析"的程序

如果说,20 世纪中叶以前,使用口语报告研究思维只是少数心理学家的偶然行为,那么 20 世纪中叶以后,西蒙等人对口语报告分析法的使用则是大规模的、系统的,他就是要通过口语报告分析了解人们问题解决内在的详细过程,以便将这一过程计算机程序化,再将其输入计算机进行模拟。所以,信息加工心理学使口语报告分析法真正有了用武之地。心理学家一方面通过口语分析获得资料,然后依靠这些资料编写计算机程序以进行模拟。我们一般都认为,口语报告分析法、计算机模拟是信息加工心理学研究思维的两种最重要方法。确切地说,口语报告分析才是其中更为基础的研究技术。

纽厄尔和西蒙从 1960 年前后开始使用口语报告分析法研究人类解决密码算术问题(cryptarithmetic problem)、逻辑问题(logic problem)和国际象棋问题(chess problem)解决的思维过程。参照纽厄尔和西蒙的实验过程,以及近期国

内类似的研究程序,我们把使用口语报告分析法研究问题解决的过程划分为以下几个阶段。

阶段一:任务设计与器材准备

任务设计(task design)就是要给被试设置一个"任务环境",考虑的因素主要包括:

第一,任务难度要与被试的平均知识水平和问题解决能力水平相匹配,或者符合实验变量本身的变化要求,既不能太难,也不能太容易;工作量的大小也要考虑研究程序的可行性和研究需要。要把自变量的变化融合到任务编排中,除考虑自变量的变化程序外,还要考虑自变量引起变化的区分度。

第二,解题过程的可展开性。利用口语分析技术研究问题解决,一定要使问题解决的过程具有可展开性,能持续较长的时间,而且解题过程容易被表达出来。

第三,问题解决行为的可展开性。有些时候,需要借助被试的外部行为表现来辅助说明问题解决的过程和机制,这就需要问题解决过程具有可观察性。有时需要被试在解题中留下较丰富的书面材料。

设计完成后,正式实验前要进行预实验,以评估任务设计是否符合要求。

器材准备。实验中需要的器材主要是录音器材(如使用小巧且录音效果好的录音笔等)。此外,还要将所需作业和完成作业需要的其他工具准备好。

阶段二:被试"出声思维"训练

通常情况下,人们解决问题时是不大出声的,除非是像课堂上老师提问时学生口头表述解题过程。因此,在实验研究程序中,正式实验前,要对被试进行"出声思维"训练。一般使用较为简单的问题作为训练材料,由主试先进行示范,然后让被试尝试进行。训练中,注意让被试学会直接口述解题思路和操作方法,而不要解释"为什么"那样做。等到被试比较习惯于"边想边说"时,就可以进入到正式实验程序。

阶段三:正式实验与口语报告材料的取得

正式实验要在安静、整洁的环境中进行。实验开始前,对被试进行言语引导,使其心理状态放松,而且能正确理解实验要求。实验开始后,被试按照要求解决问题并出声报告。在此过程中,主试要尽可能不给予被试提示和干扰,只有当其出现较长时间的停顿时,可以问"你怎么想的",或者在被试只顾将解题步骤进行下去,却忘记做口语报告时,则提示"请边做边大声报告"。

比如，纽厄尔和西蒙为了获取卡耐基大学的一名男大学生解决密码算术问题的口语报告资料，具体是这样做的：[1]被试坐在一张桌子旁，桌子上放着一枝铅笔和一张白纸；主试将实验的指示语读给被试听，并将密码算术问题以常见形式写在纸上交给他。要求被试解决这一密码算术问题，并且在整个解题过程中要大声报告他是如何想和做的，使用磁带录音机将其报告记录下来。在被试用笔在纸上写时，也不要中断录音，即使在他用笔写时没有任何口语报告材料。从这次实验中，被试的口语报告共包含了 2 186 个单词。

实验结束后，要及时将录音材料进行备份、保存。

阶段四：口语报告材料的分析

问题解决任务结束后，口语记录过程也同时结束，接下来要做的是将被试的录音材料分解为短语或短句，并用字母标记，如 B1，B2，B3……在这些编上号的句子中也包括一些主试的语句，如被试提问时主试的回答、主试提醒被试"你现在是怎样想的"，等等。在进行语句拆分时要以意群为单元，即将表达一项单一任务的短语或进行一项推论的短语作为一个独立的项目，保证每一语句项目具有一个独立和单一的语义，然后按本来的时间顺序排列。如果是属于主试的口语，则将其放在方括号内并用 E 来区分(E 代表 Experimenter)。按照这样的方法，纽厄尔等人从一个被试的密码算术题解题过程中得到了 321 个短语或短句(包括少数主试的语句)。[2]

这一口语记录材料最充分地保留了被试解决问题的真实信息。不同研究者对同一被试的口语报告进行拆分时，结果可能会出现差异，这种差异主要在于对意群识别和分割的差异，不会对被试的行为描述带来太大的影响，因为不管怎样拆分，全部的口语信息都能够得到反映。

拆分之后，就可以对口语报告材料进行定性与定量的分析，以从中发现被试的问题解决过程、使用的策略，以及在这一过程中表现出来的个人差异等。为了更直观地反映被试的解题过程，可以提取口语材料中描述出来的问题中介状态，以及被试实现这些状态转换而采用的算子，然后以时间为线索将这些中介状态和算子连接起来，构成一个问题行为图。显然，这个问题行为图恰是问题空间的

---

① Newell, A. & Simon, H. A. (Eds.) (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall, 165—166.

② Ibid., 230—244.

外部表征。

绘制问题行为图应把握的规则:①(1)一个中介状态由一个结点表示;(2)向右的水平箭头表示将某个算子应用于一种中介状态,导致另一中介状态,双箭头表示同一个算子重复应用于同一种中介状态;(3)时间流程先向右,然后向下;(4)在一个垂直方向上的多个结点代表回到同一中介状态。如图 9-10 所示,该被试从起始状态 1 开始执行算子,出现了中介状态 2,再到 3、4、5、6,其中在从状态 3 与 4 的转换中反复使用了同一个算子。出现了中介状态 6 后,没有接着对该状态操作算子,而是又返回到状态 3,操作一个算子导致了状态 7,再到状态 8,随后又两次返回状态 7 并操作不同的算子,出现状态 9 和 11 等。可见,问题行为图是根据中介状态生成时间线性排列的,能有效表征被试的问题解决过程。

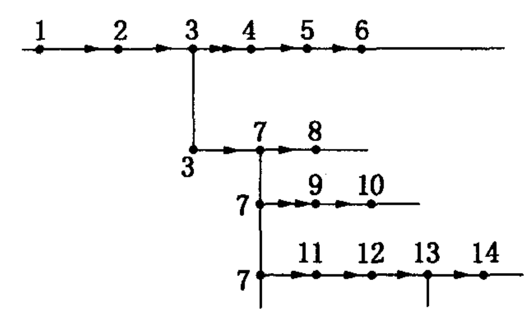


图 9-10 问题行为图片断

毫无疑问,采用口语报告分析方法,再借助问题空间概念和问题行为图绘制技术,问题解决的研究有了有效的方法,尽管这种方法的使用还显得有些复杂,但毕竟这一高级复杂心理现象领域有了可操作的研究手段。问题解决研究从此变得活跃和丰富起来了。

五、问题解决的策略与方法

"策略"(strategy)一词源自希腊语"strategos",有"计谋"之意,是指在解决问题的行为发生之前和之中所进行的行动计划。一般,人们在解决问题时可以采取多种不同的策略,而采取什么策略解决问题既依赖主体所面临问题的性质和内容,也依赖一个人已有的知识和经验的丰富性。

认知心理学把问题解决的策略划分为两大类:算法策略(algorithm strategy)、启发式策略(heuristics strategy)。

(一) 算法策略

算法策略就是按照某种确定的规则或逻辑顺序搜索问题答案的策略。这里

① 转引自:John B. Best 著,黄希庭主译:《认知心理学》,中国轻工业出版社 2000 年版,第 390 页。

包括两种问题解决的方法：第一种，对于有确定算法的问题，必须采用确定的算法按部就班地解决问题。如果在规则应用中不出错，就能够保证问题得到正确解决。第二种是尝试错误法，即把所有问题解决的可能操作一一进行尝试，从理论上说，这种方法最终总能解决问题，虽然效率不高。

但是在实际应用中，当问题解决中待尝试的可能解决方法太多时，尝试错误就是一种低级、低效的问题解决方法，而且往往由于问题解决者不能将所有可能性都进行尝试而造成问题不能得到解决。

（二）启发式策略

“启发式是指个体根据自己已有的知识经验，在问题空间进行粗略搜索来解决问题的策略。它要求以问题相关领域特定的知识为前提。启发式并不能完全保证问题解决的成功，但是运用这种方法来解决问题比较省时、省力，而且效率较高。”①

在启发式范畴内，具体的问题解决方法是多种多样的，而且不同的方法更适应于不同的问题情境，或者一个特定的问题又往往可以采取多种方法来解决。在启发式问题解决方法中，最常用的或更具有普遍性的方法是选择性探索和手段-目的分析法。

1. 选择性探索

选择性探索(selective explore)就是选择可能性最少而又获得信息最多的环节开始搜索解决问题的方法和途径。以纽厄尔密码算术问题为例(如图 9-7 所示)，有效的解决方法是：往往先从最右边一列开始探索，因为最右边一列的限制性最多，需要尝试的可能性最少(已知 $D=5$，所以 T 只有一种可能，即 $T=0$)；再看第二列，$R=2L+1$，不管 L 等于多少，R 必然是一个奇数，因 $T=0$，所以 L 不可能再为 0，那么可以推知 R 有 3、7 和 9 三种可能，这就使得需要尝试的可能性降低了许多。可见，采用选择性探索，可以使得被试需要搜索的问题空间小得多了，问题解决也因此变得容易。

2. 手段-目的分析法

手段-目的分析法(means-end analysis)是指在解决问题时，首先将问题的目标状态与初始状态进行对照，确定二者的最大差异，将消除这一差异作为问题解决的最大目标。为了达到这一最大目标，必须排除其中存在的障碍，如此层层递

① 梁宁建：《当代认知心理学》，上海教育出版社 2003 年版，第 287 页。

推,就会把问题解决的目标分解成小的目标阶梯,而采取一些小的措施和问题解决手段,实现一个一个的小目标,最终使问题得到解决。我们还可以利用海诺塔问题来说明这种方法。

如图 9-8 所示,如果要把五个圆片从 A 柱搬到 C 柱上,目标状态与起始状态的最大差异是最下边的最大圆片在 A 柱而不在 C 柱,所以要实现问题解决的最大目标就必须要将最下边的这个圆片搬到 C 柱,这是次级目标。为了实现这一次级目标,必须把上边的四个圆片搬到 B 柱上去,而要实现这一更低层次的目标,就必须把上边的三个圆片搬到 C 柱上去……如此分解下去,就可以把问题解决分解成一系列小的目标阶梯,问题的解决也显得容易了。

现实生活中,有许多问题解决都是可以进行目标分解的,它可以使许多看似困难的问题变得容易解决了。

启发式策略是问题解决者充分利用已有的知识经验,对问题结构具有一定程度的理解后,将需要搜索的问题空间减少,然后较为有效地解决问题。但是,由于这种方法不是对所有可能的问题解决方法进行尝试,所以有时会在不经意中将有效的问题解决方法先期排除,这样就不能解决问题。所以,启发式策略许多时候是有效的,但它不能保证问题一定得到解决。

最后,需要说明的是,关于推理(reasoning)和决策(decision-making)问题的实验研究虽然也有非常丰富的材料,但考虑到本课程的教学重在基本实验研究程序的掌握,不便在具体实验研究领域作太大的扩展,故将其省略。有兴趣的读者可查阅有关的《思维心理学》、《认知心理学》专著。

阅读材料 9-1

问 题 表 征

认知心理学家将信息在头脑中的呈现方式统称为表征。表征是问题解决的中心环节。国外研究者发现,问题结构上的不同并不解释问题难度上的差异,被试想象、构造或考虑问题的方式才是决定问题难度的关键。

1. 问题表征的信息加工过程

问题空间是人们对问题的内部表征,问题解决者需要利用问题本身的特征

和相关领域的知识来主动地构建它。换言之，问题解决者需要充分有效地提取问题信息，再根据问题解决者自身的经验和知识对信息进行组织和分析，发现问题结构，建构问题空间。

我们以智力数学题作为实验作业，通过详细分析34名大学生被试的问题表征过程与问题解决结果，探讨了问题表征的信息加工过程及其对问题解决结构的影响（傅小兰等，1995）。实验结果表明：(1)正确的问题表征是解决问题的必要前提，在错误的或不完整的问题空间中进行搜索，不可能求得问题的正确解；(2)正确的问题表征是对问题信息提取和理解的过程，问题规则在问题表征中起重要作用；(3)在问题表征过程中，导致建构出错误的或不完整的问题空间的主要因素包括信息遗漏（未能将问题的有关信息全部提取出来）、信息误解（对某些问题信息作出错误的分析和理解）、隐喻干扰（问题信息中潜在的歧义性使被试困惑或误导被试）等。

2. 问题规则的表征

问题表征包括四个因素：(1)问题的初始状态；(2)目标状态；(3)算子；(4)对算子的约束。傅小兰等人(1995)的研究表明，对问题规则（约束条件）的理解和掌握，在构建问题空间的过程中具有重要作用。

我们以结构简单、规则明确的换位棋作为实验作业，通过对65名大学生被试问题解决过程的详细记录，对问题规则的表征问题进行了研究（王东晖等，1997）。研究结果表明：(1)问题规则的正确表征是解决换位棋问题的必要前提。对问题规则信息的误解和遗漏是导致解题错误的重要原因。(2)发现问题的关键结构是解决换位棋问题的必要条件，应用正确的搜索方式（手段-目的分析策略、逆向搜索策略等）有助于发现问题的关键结构。

3. 专家的问题表征特点

在现实世界中，专家是指擅长于某个领域解决问题的人。我们以广告设计任务作为实验作业，通过比较广告设计专家、新手和生手的问题表征，研究了广告设计专家的问题表征特点及过程。实验结果表明，广告设计问题表征的思维结构要素主要包括信息提取、先期经验、算子和算子约束。问题表征过程分四个阶段：(1)首先从问题陈述中提取信息，专家的特点是注重提取关键信息；(2)然后进入搜索先期经验阶段，专家的特点是注重体现设计对象的具体明确的、可操作性的特征；(3)然后进入确定算子阶段，专家的特点是注重以启发式有效地缩小问题空间，选择算子时对各种具体约束的考虑已趋于自动化；(4)最后是表征

结果的输出。

——资料来源:《探索问题解决的奥秘:表征与策略》(傅小兰,2001)①

阅读材料 9-2

问题图式与学科问题表征

20 世纪 80 年代以来,认知心理学关于问题解决的研究逐渐远离纽厄尔等人提出的通用问题解决模型,走进社会生活,研究学科问题、工业问题、交通问题、就业问题、救灾问题等的解决,这些问题均处在复杂的社会背景或学科背景中,与大量专门知识相联系,属于知识丰富领域问题,而且这方面的研究也不再局限于专家与新手的比较。学生学科问题解决的研究已成为当代认知心理学研究的热点之一。最近,国内外有关数理化问题解决的研究报告越来越多,研究内容更多地集中在问题表征方面。

学科问题表征包括外部表征和内部表征,外部表征是指问题情境的成分和结构,包括物理符号、物体、维度及外部规则、约束条件或边界条件,外部表征的信息只能被知觉系统觉察、分析和加工;内部表征是问题解决者构建的关于问题的认知结构,是对觉察到的问题的初始条件、约束条件的解释。对问题的内部表征是能否成功解决问题的关键,而快速形成正确的内部表征需要大量的通过样例学习建立的问题图式为基础。

问题图式是知识丰富领域问题解决研究中的重要概念,由与问题类型有关的原理、概念、关系、规则和操作程序构成的知识综合体,具有多层含义,并与成功解决问题密不可分:(1)它是与问题解决有关的知识组块;(2)它是已有问题解决成功样例的概括和抽象;(3)可被当前问题情境的某些线索激活,进而预测或猜测某些未知觉到的线索,有助于问题表征的形成;(4)它不仅有助于问题表征的形成,而且结合了问题解决的策略、方法和程序,可以指导整个问题解决过程。因为这些缘故,研究者对问题图式的形成机制和作用机制很感兴趣,但这方面的研究成果不确定。具体的问题情境包含很多细节方面的信息,学生在解决这些

① 傅小兰:《探索问题解决的奥秘:表征与策略》,见中国心理学会编:《当代中国心理学》,人民教育出版社 2001 年版,第 37—42 页。

问题时往往被这些细节所干扰，看不到深层问题结构，所以要学习和构建问题图式，才能不断提高问题解决能力。研究认为，形成问题图式需要在具体的问题解决过程中从问题表层向深层不断排除、概括和建构："一是排除过程，从表层描述中排除不重要的细节，使得储存的信息量降低；二是概括过程，概括也会减少储存的信息，同时又会对信息进行改造；三是建构过程，建构不再是减少而是增加信息，包含对未直接表述的蕴涵信息的推断。"这些过程的反复，即不断进行样例学习，就会形成、丰富和发展问题图式。当前，要注意防止样例学习的行为主义强化训练倾向，不要忘记图式的形成是一个主体在自我意识监控下的认知建构过程。

有问题图式作为心理基础，问题表征就成为可能。拉金(Larkin)的研究认为物理问题表征的建构层次和过程是：初始表征(或"真实世界"表征)，主要由问题的表面特征决定；科学理论表征，即"物理"表征，主要由问题的深层次特征决定；数学表征。在此基础上，廖伯琴和黄希庭进一步研究大学生解决物理问题的表征层次发现，大学物理学习的优生侧重于科学理论表征，而差生受初始表征的影响，对问题的理解停留于表面特征。黄巍将有机合成问题表征分为符号表征、方法表征和机理表征，符号表征是关于问题的陈述性知识的提取与理解，方法表征是关于问题的程序性知识的提取与理解，机理性表征则是关于问题的原理和范畴的知识与结构的激活。这一研究具有重要的教育意义，他把机理表征作为最高层次和关键环节，指出在教学中要引导学生对问题的表征达到此层次，这样才能保证问题的快速解决。从不同学科问题表征结构的研究看，问题表征可分为低层次的和高层次的两大类，低层次的表征是以字面理解、表面特征的提取与加工为主，高层次的表征则是以内部结构特征和范畴理论为主。一般来说低层次表征容易受表面误导信息的影响，使问题解决发生错误。

朱新明等设计一系列的物理样例问题对学生进行训练，让他们在解决这些样例问题中掌握问题规则和条件，强化他们对条件的认知能力，提高他们对问题的表征能力，这样当他们面临新的问题要解决时就可以按照样例学习中建立的产生式对问题线索和规则进行认知和表征。这一研究虽然强调了认知作用，但还是显示出明显的行为强化训练的性质，这是该研究定位在物理符号操作范型中所必然带有的局限性。莫雷和刘丽虹则研究了样例表面内容对问题表征乃至问题解决的影响。样例所包含的信息可以分为内在原理信息与表面内容信息两个方面，内在原理信息所包含的结构或关系是解决问题的关键；表面内容则是指

问题所涉及的事物、形式、情节等具体内容，样例的表面内容对问题解决的影响也许主要表现在表征阶段，即问题表征是否按照某一样例来类比，主要是根据表面内容，当要解决的问题与激活的样例在表面和内在的内容方面都相同或同属于一个范畴时，问题将会以类比方式得到正确表征；如果二者表面内容相似或相近，但内部结构和关系不同，则样例会对问题解决产生消极影响，阻碍问题的正确表征或导致问题的错误表征。所以在教学中要提示学生解决问题时注意问题的内在内容和结构，这样才能激活恰当的样例，保证问题的快速、正确的表征。根据傅小兰等的研究，问题表征经历三个阶段：问题信息的搜索和提取、理解和内化、隐喻约束条件发现。信息的搜索和提取是一个知觉过程，这种知觉过程需要有关言语知识及言语理解能力、专门知识及问题解决经验的支持；理解和内化是对觉察到的信息的深加工，它需要知识基础、思维能力和问题解决技能；隐喻约束条件发现与意识化更需丰富知识基础、联想能力和创造能力。毫无疑问，个体的知识经验（特别是专业知识）、思维能力（特别是创造性思维能力）、问题解决的成功经验等都是影响学科问题表征的重要主体因素。在具体的问题表征中，个体难免会出现信息遗漏、信息误解和隐喻干扰等现象，因此会导致对问题表征不完整和不正确。黄巍关于化学问题解决的研究也发现，学生在专业知识总量、知识结构方面与其问题表征能力存在明显关联性，他结合国外的有关研究指出个体仅有一大堆知识是不够的，这些知识还必须不断精制化、结构化，形成互相关联的组块，才能在形成问题表征时提取这些组块，产生有利的问题表征。

——资料来源：《知识丰富领域问题表征与解决策略》（邓铸，2002）①

建议阅读文献

1. 汪安圣：《思维心理学》，华东师范大学出版社 1992 年版，第 17—34、98—112、255—284 页。
2. 邵志芳：《思维心理学》，华东师范大学出版社 2001 年版，第 15—27、50—102 页。
3. 朱新民、李亦菲：《架设人与计算机的桥梁》，湖北教育出版社 2000 年版，第

① 邓铸：《知识丰富领域问题表征与解决策略》，《宁波大学学报（教育科学版）》，2002，24(1)：32—36。

1—18、31—34、54—203 页。

4. Herbert Simon 为纪念 Allen Newell 发表的演讲。网站:*http://cognet.mit.edu/MITECS/Entry/Simon2*.

复习思考题

1. 如何理解符兹堡学派、边缘思维论、尝试错误说、顿悟说、概念形成、假设检验说、计算机模拟、口语报告分析法、问题、问题解决、问题空间、算子、问题行为图、算法策略、启发式策略、选择性探索、手段-目的分析法?
2. 简述思维心理学研究简史。
3. 概念形成实验的程序是怎样的?
4. 举例说明问题的不同分类方法。
5. 计算机模拟的基本操作程序是怎样的?
6. 使用口语报告分析法研究问题解决的一般程序是怎样的?
7. 自主查阅文献,归纳总结影响问题解决的主要内外因素。

第十章
发展与教育心理实验法

本章内容提要

发展心理学与教育心理学是我国心理学科中取得成果最为丰富的两个心理学分支领域，其实验研究的方法除包含前述各章介绍的一般方法外，还有其特殊的或典型的方法。发展心理学与教育心理学实验研究中的最大特殊性是被试的心理成熟与发展因素，它容易成为实验中的混淆变量或效度损害因子，研究中要加以平衡或控制。此外，被试流失也是这两个分支领域中需要特别注意的问题。就发展心理学来说，纵向研究设计和横断研究设计是其典型的研究模式，二者各有优缺点，结合使用可以相互弥补；皮亚杰的临床实验法是发展心理学中最具代表性的和有效的研究方法，充分体现了儿童心理学中的生态研究取向。在教育心理学中，许多实验是对教学方法、学习策略或学习方法进行效果考验，这就要对教学或教育训练的效果进行测试和比较，并涉及教育训练效果的持久性和迁移性评估问题。此外，元认知的实验研究和眼动心理学的实验研究是教育心理学中两个新的知识生长点，其实验方法尚在快速发展中。

前述各章介绍的一般研究方法均可应用于发展心理学和教育心理学的研究，而且这两个学科领域也频繁地使用着各种非实验的方法，比如观察法、访谈法、测验法、档案分析法等。需要注意的是，这两个学科领域有其特殊性，其研究目的、被试群和实验处理的方式也有其特殊性，所以实验设计也具有特殊要求，其在长期发展中逐渐形成了自己稳定的格式，如纵向研究设计、横断研究设计、阅读的眼动研究等等。本章结合大量研究实例对发展心理实验、教育心理实验的方法进行介绍，以有助于具体实验研究技能的培养。

第一节　发展与教育心理实验概述

发展与教育心理学的研究对象主要是心理能力处于快速成长中的人，包括学前儿童、小学生、中学生、大学生等。研究对象的特殊性会制约心理实验方法的具体应用，比如研究选题、实验设计中变量的分析与控制、心理成熟变量对实验过程和实验结果的影响等等。

一、发展心理学研究的特殊性

与其他心理学分支实验研究的过程相比，发展心理实验中经常会面临下列三个方面的特殊问题：研究对象的心理发展与成熟因素及其控制问题、被试的流失问题、研究效度（包括内部效度和外部效度）的保证问题。

（一）发展与成熟因素的控制

发展心理学研究中的最大特征是研究对象的心理、生理因素方面的快速发展，特别是从其出生开始到青年期结束这 30 年左右的时间内[①]，人的发展变化是快速的、多方面的、不均衡的。首先是从生理方面，个体的身高、体重、体力、脑重、性征等方面的快速发展会带来一系列的心理变化、社会角色变化，而心理和社会角色变化又会带来责任和义务的改变。比如脑重的变化，孙晔等（1982）提供的资料认为，新生儿脑重平均为 350～400 克；出生后的第一年增长最快，到年末脑重达 800～1 000 克；3 岁时达到 1 200 克。[②]朱智贤提供的资料更翔实：新生儿脑重 390 克、1 岁时 660 克、2 岁半至 3 岁时为 900～1 011 克、7 岁时 1 280 克、9 岁时 1 350 克、12 岁时接近成人 1 400 克的水平。儿童在出生时，虽然脑的细胞分化、细胞分区和分层已基本完成，脑细胞的数量也已接近成人，但是其皮质细胞结构的发展却有一个相当长的过程，一般持续到 12 岁以上。脑生理的逐步发育带来脑功能的逐步发展，特别是其认知功能的发展表现出持续性、阶段性和不均衡性。认知能力的发展又成为青少年整个心理发展的基础，包括思维能力的发展、价值观的发展等等。

---

① 吴凤岗：《青少年心理学》，北京师范大学出版社 1991 年版，第 17 页。
② 孙晔等：《出生后个体发展的心理生理学问题》，《心理科学通讯》，1982，5：42—49。

与心理发展关系密切的还有性征变化。女孩性征发展主要是从 11 岁左右开始到 20 岁左右基本完成;男孩相对较晚,一般是从 13 岁左右开始到 22 岁左右基本完成。性征发展伴随着自然的生理需求的增长、性别意识的发展,会出现一系列的心理冲突和适应问题,进而导致一系列的情绪波动。

生理和心理的发展变化既是发展心理学的实验研究任务,也经常是实验中需要控制的额外变量,因为这些因素很容易影响到实验的内部效度和外部效度。

(二) 被试流失

被试流失主要是指在实验进程中,被试由于升学、迁居、生理心理变化,甚至病亡等原因而退出实验的情况,这种情况会造成被试样本容量或构成的变化,导致实验研究效度的下降。虽然这类问题也存在于心理学其他分支的实验研究中,但可能在发展心理学中更为突出,因为发展心理学研究常常需要一个较长的研究周期。研究周期越长,被试流失的可能性就越大。

(三) 内部效度和外部效度的保证

除上述两个方面的因素会对研究效度造成威胁之外,还有其他各种各样的个体因素和社会文化因素,会给发展心理实验带来更多的变数,导致内部效度和外部效度的降低,所以在发展心理学研究中更要注意对研究效度的保证。通常可采取如下措施。

1. 采用横断研究设计缩短研究周期

即在一个时间断面上选择各年龄段的被试,在较短时间内收集较大年龄范围内的心理发展的年龄特征。但这种方法也有局限性,因为一个时间断面上的不同年龄组被试实际上是成长于不同的年代,尤其是在社会政治、经济、文化快速变革时期,所以这种方法在各种社会因素相对稳定的时期更为有效。

2. 增加样本容量,包括扩大被试取样范围

青少年的心理发展速度快,对其产生影响的因素也非常多,这就使不同环境中成长的青少年学生出现很大的个性差异,也就是说,研究对象的同质性很差。当研究对象同质性较差时,降低抽样误差的有效方法之一就是增加样本容量,或者在抽样前对研究总体进行结构分析,然后采用分层抽样或按比例分层抽样方法选取被试样本;或者扩大抽样范围,使样本包含研究对象中的各种成分,以提高样本的代表性和研究的外部效度。

3. 改变条件,重复进行实验

心理学的许多结论都不是靠单次实验得到的,而是多次重复实验之后的概

括，这样的研究可以提高研究信度，进而提高研究效度，特别是外部效度。这里所说的改变条件，主要是指改变被试群，如在不同地区重复同样的实验等。对于那些发展中受社会文化因素影响明显的心理活动来说，这种重复实验尤其重要。

二、教育心理学研究的特殊性

与心理学其他分支中的实验研究相比，教育心理实验中的最大特殊性是经常需要考察教学和教育训练的有效性，而且这种有效性也包括教育训练的可迁移性。因此，我们这里讨论教育心理实验中关于教育训练的有效性检验，包括可迁移性检验，以及内部效度和外部效度的保证问题。

（一）教育训练有效性的检验

在教育心理学中，有许多研究涉及教育教学有效性的问题，这样的研究首先面临的一个问题就是如何评估教育训练的有效性。有效性评估包括近期效应和长远效应的评估：近期效应评估主要是教育训练实施之后，立即进行的效应测量；长远效应评估则是在教育训练实施之后，间隔一定的时间进行一次或多次效应评估的方法。教育教学过程中，有许多教育训练追求的是长远效应，而不是短期效应，这时就要设计在训练之后的不同时间周期进行效应测查，而且这种测查要比较巧妙，有时需要采用间接方法，以避免测查材料的学习效果带来的对训练效应的高估现象。

此外，教育训练希望训练的效果具有可迁移性，即对跨情境下的学生的行为和心理发展能产生积极影响，这就要在实验处理或教育训练实施之后，立即或延迟一段时间后，改变刺激情境，直接或间接地测查实验处理或教育训练的效应。

（二）内部效度和外部效度的保证

1. 发展与成熟因素的控制

有一些教育心理实验也像发展心理实验一样，需延续较长时间，如一学期，一学年，甚至更长，那么伴随实验处理或教育训练，就可能存在心理或生理成熟的系统性变化，这些变化也就成为混淆因子，影响研究的内部效度。对此可采取一些有效的方法加以克服，比如在实验设计本身许可的情况下，缩短实验周期，或者选择那些心理发展速度较慢的年龄段的被试参加实验。

2. 取样偏差的控制

教育心理实验中，被试的选取一般都难以做到随机取样，更多时候是采用便

利抽样或整群抽样，因此使用有偏样本进行实验是常有的，这就对研究的外部效度造成了威胁。为了控制取样偏差对研究外部效度的影响，可以采用较大规模的教育心理实验，即增加实验被试量，并且在选取被试时尽量考虑到学生的地区分布、层次分布、年龄分布、性别比例等等，特别是要考虑到所选实验班级中学生的智力水平等等。

3. 实验者效应的控制

在教育心理实验中，还经常存在实验者效应问题。因为许多教育心理实验是与教学过程紧密结合的，实验处理由任课教师来完成。这些任课教师存在很大的个体差异：年龄、性别、受教育年限、教龄、教学风格、亲和力等等，这些差异可以带来实验者个人因素与实验处理之间的交互作用。为此，要在实验中尽量控制实验教师的上述特征，尽量减少实验者因素给实验结果带来的偏差。

三、发展与教育心理实验设计的要求

根据上述分析，发展与教育心理实验有其特殊性，在研究设计上也有其特殊要求。从近年来关于心理学方法论的讨论看到，发展与教育心理实验强调具有更多的生态性，有更明显的现场研究的特征①，所以研究中更多地使用控制条件下或干预条件下的现场观察。

（一）实验干预下的观察

由于儿童心理活动有突出的外显性，通过观察其外部行为可以了解他们的心理活动，因此观察法是发展心理学研究最基本、最普遍的一种方法。发展心理学早期的许多研究大多采用这种方法。例如陈鹤琴的《儿童心理之研究》主要是通过观察收集资料的。在教育心理学中有经验的教师经常在实际工作中对学生做大量的观察，捕捉学生的行为表现，从而采取恰当的教育手段。但是不加控制或干预的观察，难以保证研究的效度和效率，研究者会比较被动。采用折中手段，则是在观察中加入干预因素，即进行实验干预下的观察。看来，心理实验研究也少不了利用观察技术。观察技术的使用通常包含三个步骤：

第一，确定观察内容。如研究教师教学态度对师生交往及学生学业成绩的影响就需要考虑在哪个学校、哪个年级和班级进行，要观察哪些现象，如何在较

① 林崇德：《发展心理学》，人民教育出版社 1995 年版，第 77—88 页。

为自然的状态下对师生交往行为作客观记录等。

第二，选择观察策略。常用的观察策略有参与观察策略、取样观察策略和行为检查表策略等。

第三，制定观察记录表。目前在制定观察记录表时通常采用观察代码系统，它们是为观察、记录和随后分析处理的方便而制定出的一些符号代码系统。

在运用干预下的观察研究心理发展问题时需要注意以下几点：

第一，对要观察的问题有基本的了解，观察目的要明确。如研究师生互动行为时，观察者要深入课堂，观察课堂上教师提问、学生举手与发言、教师板书与面部表情等。

第二，尽量使被试自然放松，处于正常活动状态，尽量避免他们意识到被观察。观察者可以借助单向玻璃、电视、潜视镜等进行观察，这样可以减少被试的行为表现受到的影响。

第三，要善于记录与观察目的有关的事实，以便事后进行整理分析并提出进一步研究的意见。例如，研究学前儿童言语发展，由于学前儿童言语表达方式与成人不同，所以应避免使用成人语言作记录。为提高记录的准确性，可使用小录音机、录像机等器材。

第四，除观察被试的一般言行外，还应分析被试的相关材料，如作文、日记、各种作业等，以帮助研究者对观察到的被试行为作更准确的解释。

观察有其独特的优点。因为它是在被试行为发生的当时现场进行观察、记录的，能够收集到比口头报告或问卷调查更客观、全面、准确的资料，所以在发展与教育心理实验中多采用一些观察技术，可以提高研究的外部效度，提高研究结论的可靠性。

（二）实验设计的具体要求

实验法是指对研究的某些变量进行操纵和控制，创设一定情境，以探讨被试心理发展的原因和规律的研究方法，包括自然实验法和实验室实验法两种。发展与教育心理研究中更多的是使用自然实验法。

采用自然实验法研究发展与教育心理，首先要确定研究的课题和实验的目的，然后充分地估计和分析实验进程中可能出现的变量，确定实验的实施方案，控制其中各种不利于实验目的实现的因素。比如，在发展心理学中，对儿童心理进行实验研究时应注意以下几点：

第一，实验目的、材料和方法都应该与教育的原则相适应，有助于儿童身心

发展；

第二，儿童心理实验室的布置，应尽量接近儿童的日常生活、学习环境，使儿童在实验条件下有较为自然的行为表现；

第三，实验进行中应考虑到儿童的生理状态、情绪表达与控制能力，使其能积极参与实验活动。

自然实验法的特点在于它能使实验情境保持较自然的状态，同时又能按照实验目的对某些条件加以有计划地操纵或控制，因而能较好地保证研究具有较高的内部效度和外部效度。当然，也正是由于自然实验法更接近自然，自然环境的复杂性也给研究实施带来各种困难。

在实验室实验中，研究者能够严格控制和操纵变量，有效地揭示变量间的因果关系，但由于实验室实验研究往往需要使用精密的仪器设备控制实验进程，使研究情境与被试的实际生活情境出现较大距离。如幼儿在实验室中很容易感到紧张和害怕。还有，在一些品德心理学研究中，被试在有明显控制的实验情境中很容易掩饰自己的内部动机和信念，使研究者收集到存在虚假成分的资料，导致结论出现偏差。但有时实验室研究是必需的，这就需要采用有效方法，将实验室情境布置得较为自然一些，使研究进程能够得到有效控制的同时，又能保证研究有较高的外部效度。

当前，研究者大量采用了以计算机控制实验的进程。在这样的实验中，额外变量一般容易得到控制，数据的记录和整理也比较精确，这是心理学实验技术发展的一个重要趋势。如数学教学心理、阅读心理、问题解决的模拟等都可用计算机屏幕显示刺激材料，记录反应的正确与否、速度及解题过程等。

第二节　发展心理实验设计

发展心理实验研究的设计有两种典型模式：纵向研究设计、横断研究设计。

一、纵向研究设计

纵向研究设计是指在一定时间周期内，对同一组被试的一个或多个特定变量作两次或更多次测量以调查其发展的年龄效应。纵向研究设计是发展心理学研究方法中的一个特色，现代儿童心理学的奠基人普莱尔（Preyer，W. T.）最早

运用纵向研究技术系统考察了儿童的心理发展。他每天三次对自己的孩子进行观察,并辅以简单的实验技术,历时三年,记录了大量有关儿童心理发展的资料,随后撰写的研究报告被称为"儿童传记"。普莱尔在纵向研究的基础上,完成并发表了他的《儿童心理》(1882)专著,此书成为世界上第一部儿童心理学领域的经典著作。纵向研究设计的基本模式如图 10-1 所示。①

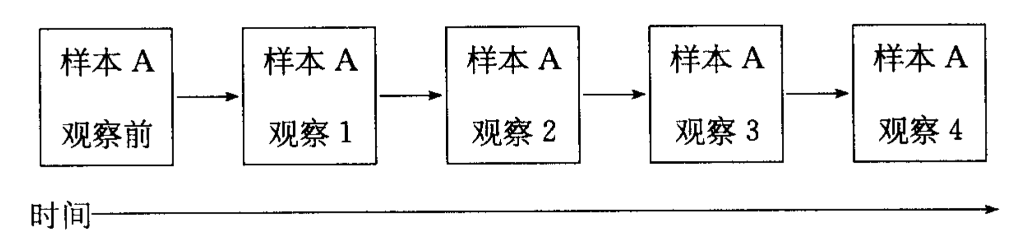


图 10-1　纵向研究设计示意图

例如,用纵向研究设计考察时间价值观与年龄的关系。研究者可以测量一组被试在 15 岁时的时间价值观,然后分别在 5 年后、10 年后再测量这组被试的时间价值观,这样就可以得到该组被试在 15 岁时、20 岁时、25 岁时的时间价值观。

纵向研究设计与时间序列设计非常相似,不过在纵向研究设计中,不施加实验处理,研究的自变量或准自变量就是时间或年龄。纵向研究可被描述为先后对一组被试进行的一系列观测,其中首次观测与最后一次观测的结果的差异应被看作是发展效应。

纵向研究设计和时间序列设计的区别并不总是清晰的。例如,考察父母离婚前后的一段日子里一组青少年的幸福感。这可看作是纵向研究,因为它考察一个被试组在较长时间里的变化。但是,它也可以被看作是时间序列研究,其过程是把事件(离婚)发生前的一系列观测与事件发生后的一系列观测进行比较。

国内心理学工作者张文新等(2003)采用纵向研究设计研究了学前儿童攻击行为的发展变化。②研究采用整群抽样方法,从济南市四所幼儿园的小班选取 7 个班的 217 名幼儿为被试,然后分三次对这些儿童的攻击行为进行观察,每一次观察都持续两个月。三次观察的时间是:第一次观察的时间是 2000 年 10～11 月,第二次观察的时间是 2001 年 4～5 月,第三次观察的时间是 2001 年 10～11

---

① 王重鸣:《心理学研究方法》,人民教育出版社 1990 年版,第 55 页。

② 张文新、纪林芹:《3～4 岁儿童攻击行为发展的追踪研究》,《心理科学》,2003, 26(1):49—52。

月。不过,因为转园等原因,到第三次观察时有效被试只有163名。为了使观察中的记录比较客观和可靠,研究者将儿童的攻击行为界定为有意伤害他人、损坏或抢夺他人物品的行为。按照实施方式,儿童的攻击行为又可以划分为身体攻击、言语攻击和间接攻击三种形式。其中,身体攻击包括打、扭、拧、指戳、砸、挤压、踢、推搡、碰撞、抓、咬他人,打掉、损坏、抢夺他人物品,强占座位空间等;言语攻击包括说脏话、骂人、取笑、嘲讽、叫取外号等。间接攻击又称心理攻击或关系攻击,是通过第三方实施的攻击行为,包括散布坏话造谣、唆使打人、游戏活动排斥等。

张文新等的研究发现,儿童最普遍的攻击形式是身体攻击,其他两种攻击行为发生频率较低;大多数攻击行为属于主动性攻击和工具性攻击;男孩的攻击行为总体上多于女孩,但女孩的间接攻击多于男孩。

二、横断研究设计

横断研究设计是指在同一时期同时测量处于各不同年龄段的被试变量,然后通过比较得到某一个或某几个被试变量与年龄的关系,这里的每一被试组代表着一个不同的年龄段。例如,国内有一项涉及23个省市在校青少年学生的思维发展的研究,就属于横断研究。研究者在同一时间对来自23个省市、自治区的不同类型学校(包括城市一般学校、省市自治区及重点中学、农村中学)的初一、初二和高三学生在形式推理、形式逻辑法则和辩证逻辑思维的发展水平进行了测查和比较。①

横断研究设计可看作是非等组设计,不同的被试组不是通过操纵自变量产生的,而是通过现成的被试变量(年龄)确定的。它的一个明显优点是研究者无须等待许多年,而是可以在一天中完成对多年成长效应的考察。例如,考察时间价值观与年龄关系的研究,可以采用前述的持续10年的纵向研究,也可以采用在半天中就能完成的横断研究设计,即分别在15岁、20岁和25岁的人群中各抽取一定量的被试组成三个被试组,要求他们同时完成一项关于时间价值观的测量,就可以在半天当中完成这一研究所需资料的收集。横断研究可在短时间内收集数据,不需要与被试进行长期合作,因此研究者也不需要付出很长时间

① 王重鸣:《心理学研究方法》,人民教育出版社1990年版,第56页。

和代价对被试进行追踪,这是它的明显优点。

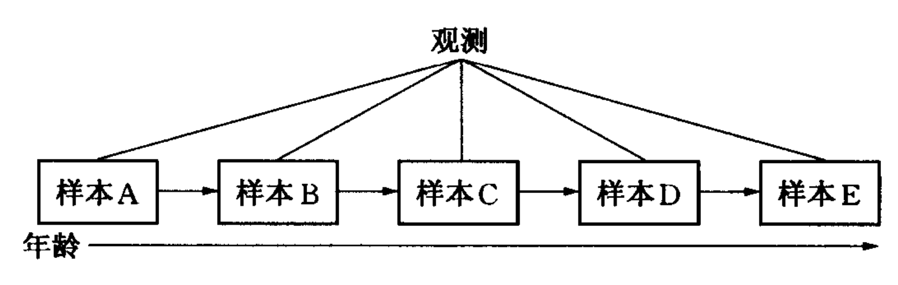


图 10-2 横断研究设计示意图

横断研究也存在明显缺点。首先,研究者无法说明各组被试是如何成长的,因为该类研究未对被试作追踪研究;其次,研究中可能会存在许多难以控制的其他机体变量,导致研究变量的混淆,最典型的混淆因子就是不同年龄的被试成长的年代存在明显差异。例如,30 岁的妇女不仅是比 50 岁的妇女年轻,而且她们成长的环境可能很不同。对这两组妇女而言,受教育的机会、工作年限和社会期望都不同。年龄相同且生活在相似环境中的被试叫做同辈,但年龄不同的被试是生活在不同环境中的,因此造成的差异叫做代间效应或辈间效应。看来,在横向研究中,组与组之间的差异未必都是年龄带来的差异,这里存在明显的混淆因子——成长的不同时代或环境。

这里有一个很好的例子可以说明辈间效应对横断研究结果的影响,即智力与年龄的关系研究(Horn, 1978)。有许多研究表明人在 20 岁至 50 岁之间智力呈下降趋势,但同时也有许多研究得到不同的结论,认为 20 岁至 50 岁之间智力呈轻微的下降趋势或无下降趋势。为什么出现两种不同的结论呢?原因在于研究的设计类型不同。表明智力随年龄增长而下降的结论一般是从横断研究中获得的,横断研究的结论可能受到辈间效应的影响,其使用的被试组之间不仅存在年龄上的差异,而且因为他们生活在不同的年代,也因此会有差异,这可能正是他们智力分数不同的原因所在。而那些表明人的智力水平稳定的研究则多是属于纵向研究,即在较长的一段时间内对同一组被试进行的追踪研究。

友谊作为一种亲密持久的同侪关系被认为是在青年初期才发展起来的,但方富熹等(1993)采用横断研究设计探讨了小学低年级儿童(7、9 岁)对友谊关系的认知发展,并将研究结果与国外同行的研究进行了比较。①他们发现,儿童对

① 方富熹、方格、王文忠:《小学低年级儿童对友谊关系的认知发展》,《心理学报》,1993, 25(1):1—7。

友谊关系的认知发展既有跨文化的普遍性,也存在文化差异。随后,他们对12岁和15岁儿童的友谊关系发展作了进一步研究①,试图揭示我国儿童的友谊关系从不成熟到成熟的发展过程及其主要影响因素。研究中,他们选取了7、9、12、15岁四个年龄段的被试组,其中7岁组被试72人,其余三个组各90人,四个年龄段被试分别来自小学一、三、五年级和初中二年级的学生。为了控制取样误差,研究者抽取的被试来自下列3类不同的学校:北京市区重点中小学各两所、市区普通中小学各三所和郊区农村中小学各一所。每类学校7岁组被试各24人,其他年龄组各30人。

研究采用个别访谈方式进行。开始时,先由主试叙述一个以承诺为主题的两难故事:主人公已答应老朋友的邀请,约定到老朋友家谈心,这时又接到一新结交的孩子的邀请,要他一起去看一场精彩的电影。这两件事正好出现了时间冲突,主人公不知如何是好。故事叙述完后,主试向被试提出了一系列开放式问题,要求他们根据自己的理解或观点回答。研究中提出的问题主要包括:(1)对问题情境的定义:如主人公碰到了一个什么问题,这为什么成为问题?(2)决定选择:如主人公会去谁那儿呢,为什么?(3)对后果的考虑:让被试设想主人公在作出这种或那种行为选择时的想法和心情,以及久等主人公而主人公失约不到时老朋友的心情;(4)问题的解决:让被试提出建议帮助主人公解决好这一矛盾;(5)对决定的评价:让被试指出主人公应该到谁那儿去,为什么?这里的1~4类问题在于探察被试的观点采择能力,第5类问题则属于探察被试道德判断能力的发展水平。此外,主试还提出其他一些问题以深入了解被试对友谊、承诺等概念的理解。研究结果表明:儿童在有关友谊关系及其认知方面的发展受到其面临两难问题时选择能力的影响,也受到其对涉及友谊关系的道德规范的理解的影响,表现出阶段性的发展模式。

在发展心理学研究中,横断研究比纵向研究更常见,因为它能在较短时间内获得年龄因素对心理发展的影响。比如,王美芳等采用横断研究设计对6~12岁儿童嵌套思维的发展进行了研究②,其数据收集在一个月内完成。嵌套思维是指对他人思想的认知,可以分为一级嵌套思维、二级嵌套思维等。一级嵌套思

---

① 方富熹、方格、凯勒:《对友谊关系社会认知发展的跨文化比较研究》,《心理学报》,1994, 26(1): 44—50。

② 王美芳、张文新、林崇德:《6~12岁儿童嵌套思维的发展研究》,《心理科学》,2001, 24(4):492—493。

维的形式如"我认为张天知道球是王雷拣的",二级嵌套思维的形式如"我认为王雷认为李瑞华知道球是王雷拣的"。研究者采用分层随机抽样方法分别从济南市一幼儿园大班和某小学的一年级、三年级、五年级各抽取16名儿童,其年龄分别为6岁+3个月、8岁+3个月、10岁+3个月、2岁+3个月。施测均在5月份进行。研究结果表明:6岁儿童的一级嵌套思维已基本成熟,二级嵌套思维也有了一定程度的发展,6～8岁是儿童二级嵌套思维的快速发展时期。

三、皮亚杰的临床实验法

毫无疑问,皮亚杰是迄今为止发展心理学领域最杰出的科学家,他关于儿童认知发展的研究方法和研究成果都令人赞叹。单就其研究方法来说,既不是测验法,但也采用测验技术;既不是实验法,但也采用实验技术;既不是观察法,但又常常使用观察技术。他将多种技术应用于自然情境下儿童表现出来的认知水平的记录与分析,充分体现了发展心理学研究中的生态性,所以一般将皮亚杰的研究方法称为临床法。①"一定意义上,临床法也是一种实验方法,因为研究者在展开一项研究之前,总是事前对研究的主题有所准备了,包括作出某些假设。但重要的是临床法的所谓实验并不施以严格的变量控制,因而可以避免一般实验法可能有的'系统误差'。"②我们就将其方法称为"临床实验法",因为在皮亚杰的许多"临床"研究中其实是包含"变量操纵"的,如"三山实验"。

研究实施中,皮亚杰根据研究内容和研究对象的不同,既采用了纵向研究设计,也采用了横断研究设计。"婴儿研究中主要用纵向法,作长期跟踪研究;在幼儿期,则主要采用横断法,在较短时间内同时对不同年龄儿童加以研究。"③

皮亚杰主要通过设置某种游戏情境,然后与儿童进行灵活的个别谈话来开展研究。皮亚杰认为自己的研究方法与富有经验的临床医生诊断与治疗情绪问题的方法很相似,其实质在于它的灵活性——研究者可以根据需要随时调整预定的研究程序,从而灵活地运用各种未经预先规定的方法来探究儿童的反应。

① 郭力平:《学前儿童心理发展研究方法》,上海教育出版社2002年版,第239—243页。
② 李其维:《破解"智慧胚胎学"之谜——皮亚杰的发生认识论》,湖北教育出版社1999年版,第122页。
③ 郭力平:《学前儿童心理发展研究方法》,上海教育出版社2002年版,第239页。

皮亚杰的研究具体是如何进行的呢？我们以他在研究相对空间概念的理解力时所做的"三山实验"为例来说明：他制作了三座山的模型，将这些模型摆放在一张桌子上，这三座山的颜色有所不同，而且在第一座山上有间房屋，第二座山顶上有一个红色的十字架，第三座山上覆盖着白雪。如果从前后左右等不同角度来看这些模型，则会形成各不相同的空间视觉效果。研究者先让儿童围着模型绕行一周，从不同角度知觉这些模型，然后坐在模型一侧。主试指指坐在模型另一侧的小朋友，问被试："那个小朋友看到的山是什么样子的？"被试很难回答。接着给被试出示从不同角度拍摄的"三座山"的照片，让被试挑出另一侧小朋友看到的山是哪张照片的样子。研究的结果发现，8 岁以下儿童不能正确地完成上述任务，大多数 6 岁以下儿童是按照自己知觉的模式进行反应。由此，皮亚杰发现了幼儿在对事物进行判断时是以自我为中心的，尚不能站在"他人的立场"来理解事物。

临床实验法看似简单，实际上对研究者提出了更高的素质要求。皮亚杰在《儿童关于世界的概念》一书中，对临床法作出了清楚的表达：

> "我们认为，在儿童心理学领域使用临床法，正如心理病理学一样，初学者需要至少一年的日常实践，才能度过其杂乱摸索的阶段。在对孩子提问时，不能说得太多，也不能含有暗示，这是很难做到的。尤其困难的是，要在预先考虑好的系统观念，与由于缺乏具体假设而致的杂乱无章之间，找出一条中庸之道。一个优秀的研究者必须具备两种似不相容的品质：他必须知道怎样观察，也就是说，让儿童自由自在的讲话，使他们无所顾忌，毫不犹豫；同时，他又必须善于时刻警觉地捕捉有关的要素，在他的头脑里必须时时刻刻存在着所要探讨的某种假设或某种理论，不管它们是对的还是错的。鉴于临床法运用时的诸多困难，新手应接受一定的训练。"①

四、纵向研究与横断研究的比较

纵向研究与横断研究是发展心理学领域最重要的两种研究模式，在教育心理学中也发挥着非常重要的作用。这两种方法各有特点，也各有优缺点。

---

① 转引自：郭力平：《学前儿童心理发展研究方法》，上海教育出版社 2002 年版，第 240—241 页。

研究者可以根据研究课题的性质、研究对象的年龄特征、研究条件的限制等作出选择。

纵向研究的优点是，能对同一个、同一组或几组被试作长期的跟踪研究，系统地积累年龄因素在其心理发展中所起的作用，特别是自始至终参与研究的所有被试是同一批被试，这就不存在辈间效应带来的变量混淆。但是纵向研究也存在明显缺陷：首先，非常耗费时间，比如有些研究要持续几个月、几年甚至几十年。其次，纵向研究在操作上也要付出很大代价，研究者不仅需要对被试进行追踪，必要时还要费力地劝说他们重新回来参加研究。如果研究历时很多年，还要不断地训练主试。再次，这些研究也受被试高退出率的影响，尤其是当退出的被试与留下来的被试有系统差异时，研究结束时的组与研究开始时的组可能有不同的特点。最后，纵向研究要对同一组被试进行多次重复观测，研究后期获得的分数必然或多或少地受到先前测量的影响。

横断研究的优点是它在一个较短的时间内，观测各种年龄段的被试，以比较年龄因素对个体心理发展的影响。这种研究的效率比较高，也由于能在较短时间内完成，所以就可以节省实验控制的费用，实验过程也少受被试流失的影响。横断研究的最大缺点是存在辈间效应，也不能有效地评估一个被试的发展历程。此外，横断研究中还存在实验材料设计的难点，因为同时观测的被试处在不同的年龄段，心理发展水平不同，这时使用同样的测验材料势必会出现“太难”或“太容易”的情况；如果对不同年龄的被试使用不同的材料，则必然会引入新的混淆变量。因此，横断研究和纵向研究都不是尽善尽美的研究技术，各自的优点和缺点都很突出。

我们可以将上述分析总结如表 10-1 所示的内容。

表 10-1　纵向研究与横断研究的优缺点

	纵向研究	横断研究
优点	无辈间或代间效应 能系统评价被试的行为变化	资料收集的效率高 研究控制成本降低 少受被试流失的影响
缺点	耗费时间 被试流失带来混淆变量 研究控制成本提高 存在练习效应	存在辈间效应等混淆变量 不能系统评价被试的行为变化 测验材料的设计较易引入混淆变量

第三节　教育心理实验设计

教育心理的实验研究选题很广泛，比如关于学习动机的研究、概念形成的研究、推理机制的研究、问题解决的研究、学习策略的研究、教学与学习风格的研究、认知方式的研究、元认知研究和眼动研究等。前述的许多方法都可应用到这些选题中，研究设计本身也没有太多特殊之处。相比而言，下述三个方面尚需作一些专门讨论，因为这些方法较为特殊或在教育心理学领域具有典型性。

一、实验组与控制组的使用

实验组和控制组的比较是心理实验中最通用的形式，它是前文介绍过的组间设计的另一种表达，只不过许多教育心理研究者习惯于设置一个不接受任何处理的控制组，就使得这种设计成为一种典型形式。这一实验形式更多地被用于教学策略或教学方法的研究，基本的研究程序主要由以下步骤组成。

第一步：提出研究课题。如考察某种新的教育训练方法、教学方法的效果。

第二步：编制研究方案和相关材料。这里所说的材料往往是一些学习绩效的评估材料。

第三步：选取实验被试。在教育心理实验中，多采取整群随机抽样或便利抽样的方法抽取被试，即多使用学校中现成的班级作为实验的被试组。在选取被试组时，一般至少需要两个组，其中一个组是实验组，另一个组是控制组。当需要对两种以上的教育训练进行考验时则需要 3 个以上的被试组。实验组就是预计要接受实验处理的被试组，控制组就是不让其接受实验处理的被试组。控制组为实验观测提供了一个基线水平，因此也叫做对照组。

第四步：实验处理的实施。在一定的时间周期内，对实验组被试施加实验处理，而控制组被试未接受任何新的处理。所有被试组除在接受或不接受实验处理方面的差异之外，其他有关影响因素会得到一定的控制，尽量保持相同水平，主要如作息时间、授课教师、班级气氛等。

第五步：实验处理后的观测。在预定的实验处理结束时，同时对实验组和控制组进行相同的测试，以得到所有被试在完成同样作业时的分数。

第六步：实验处理效应的检验。观测结束后，对实验组与控制组被试的测试

分数进行独立组的 t 检验或方差分析，以考察实验处理是否产生了积极效应。

在这一研究过程中，我们一般会把实验组和控制组看作是相等组。不过，由于这里经常采用的整群抽样和便利抽样，真正得到相等组的可能性较小。如果被试组之间存在较为明显的差异，实验结果的处理方式就变得不可靠啦，这时一般要采用准实验的研究程序。

在准实验研究程序中，因为考虑到实验组和控制组可能存在一定的差异性，控制组的观测成绩就不能作为实验组的基线水平，所以准实验研究需加上前测程序。在实验处理实施前，使用相同的测量工具对实验组和控制组被试进行测量，这样就可以得到前测成绩。组间前测成绩的差异量可作为后测成绩比较的校正值，或者在实验处理结束后，分别计算实验组前测-后测成绩的差异量、控制组前测-后测成绩的差异量，最后采用 t 检验或方差分析对这些前测-后测成绩的差异量进行比较。

二、元认知研究的实验方法

元认知(metacognition)是 20 世纪 70 年代出现的新课题，曾被美国心理学家伯尼奥(Boneau, 1990)看作是认知心理学和发展心理学两个领域中的"前沿百题"(top 100 topics)之一，它的最简洁和最具有概括力的定义是"关于自己认知的认知"，是"跳出一个系统后去观察这个系统"的认知加工(Hofstadter, 1979)。弗拉维尔认为(Flavell, 1978)，"元认知是个人对自己的思维和学习活动的认知和监控"，"从更为广泛的意义上，元认知被定义为任何以认知过程与结果为对象的知识，或是任何调节认知过程的认知活动。它被称为元认知，因为其核心意义是对认知的认知"。从近 30 年积累的文献看，有关元认知的定义基本上没能脱离弗拉维尔提供的框架。在弗拉维尔的元认知的定义中至少包含下列几层含义：(1)元认知是指向认知主体认知活动的；(2)元认知的内容是指关于"我"的认知，而不是"我"的别的方面，具体包括对"我"的感知、记忆、思维、言语及相应状态等的认知和体验；(3)元认知还包括对自身认知能力的认知，而且这种认知会影响到个人对认知过程的调控。

元认知研究的任务一般包括元认知的结构与功能、元认知能力的发展性评估、元认知能力的教育训练、元认知发展的影响因素等等，这其中包含的元认知监控是研究的核心。研究方法主要是测验法和实验法，其中实验法主要被用于元认

知教育训练的研究方面，所以本章只对元认知教育训练的实验研究进行介绍。

（一）元认知训练的有效性

元认知训练是否有效，主要考察这种教育训练是否增进了学习者的策略知识、监控意识和监控绩效，而且在对这些方面进行评估时，还要看其效果的持久性和迁移性。

1. 训练效果的持久性

关于元认知训练的研究主要集中在提高儿童认知监测的主动性、精确性及使用各种认知策略的有效性方面，涉及的内容包括记忆、阅读理解、写作和解决数学问题等过程中的认知监测和策略使用，尤以记忆策略的训练为最多（陈英和，1996）。研究中一般认为元认知训练或多或少会产生一些积极效果，但是这些效果是否具有持久性却是很关键的，因为只有持久的效果才算是真正有价值和真正影响到了被训练者的认知结构。所谓训练效果的持久性是指训练过的元认知行为能长时间地保持。布朗和巴克利（Brown & Barclay, 1976）考察了元记忆训练对儿童回忆成绩预测及实际成绩的影响①，他们指导轻度学习障碍儿童在项目学习中使用自我检查策略，对学习效果不断进行自我反馈，然后分别在训练后1大、2周和1年后进行回忆准备性的估计和实际的回忆测验，结果这些儿童比心理年龄相当但未经指导的控制组儿童的成绩好得多。布朗（Brown, 1978）在研究中对儿童估计自己回忆成绩的能力进行训练②，因为他认为如果儿童不知道自己记忆能力的局限性，也就不能期望对他们进行指导以改善其记忆。他把两个年龄组的儿童各分成两个组，一个组参加反馈练习，另一个组则不参加反馈练习。反馈练习就是每次预测和回忆之后，将其预测和回忆结果反馈给被试。训练结束后，分别在训练后1天、2周和1年后对所有被试进行成绩预测的和实际的测验，结果年龄较小的儿童只在训练后一天的测验中有一些提高；年龄较大的儿童在训练后一天的测验中有很大提高，而且在经过一段时间后仍能保持得较好，证明了元认知训练的效果是可以较长时间保持的。

当然，不同研究者的结论很不同。有的训练可以产生持久和广泛的影响，但

① Brown, A. L. & Barclay, C. R. (1976). The effects of training specific mnemonics on metamnemonic efficiency of retarded children. *Child Development*, 47, 71—80.

② Brown, A. L. (1978). Knowing when, where, how to remember: a problem of metacognition. In R. Glaster (Ed.), *Advances in instructional psychology*. Hillsdale, HJ: Erlbaum. 367—460.

更多的研究认为元认知训练的效果还依赖训练情境,即只在与训练相同或相近的认知情境中,训练的效果才会显现出来,表现为一种"黏着"现象。这说明元认知训练还受诸多因素的制约,在一定条件制约下,儿童只学会了对策略的具体操作,但对策略的功能缺乏了解或理解,即关于认知策略的元认知水平没有得到相应的提高。"因此,策略训练的核心任务和最终目标还是要提高儿童的元认知水平"(陈英和,1996),一些研究者在此方面也进行了有益的尝试。

2. 训练效果的迁移性

初期的研究(Brown & Campione, 1977)认为小学高年级学生已能将一种训练情境中学会的元认知技能迁移到新的认知情境中。①但布朗等人(Brown et al., 1978)评论说,这些训练的结果并不能看成是元认知训练迁移的证据,因为在这些研究中只是训练的材料与迁移测验的材料有所不同,而其他如实验者、实验环境等因素没有改变,这可能成为迁移的促进因素。兰格和皮尔斯(Lange & Pierce, 1992)在方法上克服了上述缺点后发现,即使给予学前儿童很简短的元认知指导②,即使在测验阶段使用明显不同的刺激材料和实验者,他们也能在训练后使用学习过的组织策略。如何才能提高元认知训练的可迁移性呢?这里介绍两个比较有效的训练实验程序。

德希勒和舒马克(Deshler & Schumaker, 1988)的训练实验程序:(1)与学生进行某种约定,要求他们将课堂上学到的有关策略尽可能运用到平常的各种作业中;(2)对于改进哪一种策略与提高哪一门学科的成绩,学生自已有权决定;(3)采用灵活多变的方式,给学生提供一些与原来的训练内容有所区别的任务,让其进行练习;(4)教给学生如何自学这些策略以及如何自我检查和奖励;(5)在教学活动一开始就教给学生应在何时何地使用这些策略;(6)对学生的迁移给予及时强化。

普雷斯里(Pressley, 1991)的训练实验程序③:(1)一次只向学生介绍某一种

① Brown, A. L., Campione, J. C. & Murphy, M. D. (1977). Maintenance and generalization of trained metamnemonic awareness in educable retarded children. *Journal of Experimental Child Psychology*, 24, 191—211.

② Lange, G. & Pierce, S. (1992). Memory-strategy learning and maintenance in preschool children. *Developmental Psychology*, 15, 543—551.

③ Pressley, M. (1991). Can learning disabled children become good information processors? How can we find out? In L. Feagans, E. J. Short & L. Meltzer (Eds.), *Subtypes of learning disabilities: theoretical perspectives and research*. Hillsdale, NJ: Erlbaum, 137—162.

策略；(2)在一系列的练习中，教师要反复向学生示范一种新策略的使用方法；(3)教师要根据每个学生的具体情况进行个别辅导，然后有重点地向全体学生做解释说明；(4)教给学生对自己认知活动的进程进行自我监控的方法；(5)教给学生如何掌握运用策略的时机，并让他在广泛情境中练习使用这些策略。

(二) 元认知监控能力的训练

洛迪科等人(Lodico et al.，1983)进行的一项训练试验分为以下三个阶段。①

第一阶段：对实验组儿童进行游戏活动中策略监测的训练。首先告诉儿童有很多种玩游戏的方法，而且有些方法是更有效的。为了游戏做得更好，他们必须选择更好的方法。接着让他们参加两项游戏：一是用手画一个圆，再用圆形饼干切割刀画一个圆，随后问他哪一次画的更好，为什么？接着问如果再让他画一个圆而且他又想画一个最好的，他将怎样做？二是要求儿童记住一个字母表，在他们学习完一个随机排列的字母表后进行检测，再要求他们重新排列字母顺序并读出名称(每个儿童学习的字母表都不同)，然后进行第二次测验，随后问他们哪一次记的更好，为什么？并问如果再让他们做这个游戏，他们将如何做？在这一阶段要不断地把结果反馈给儿童，并让那些不能回答提问的儿童重复游戏。与此同时，控制组儿童也在做同样的任务，学习同样的策略，但是不进行指导，也不提供反馈，然后告知还有更多的游戏要做。

第二阶段：让实验组和控制组儿童参加同样的有策略的学习与回忆实验共四次，分别为：(1)用复述方法学习并回忆配对词表，即要求被试在学习 10 对词表时，将每对词放在一起来记；(2)用造句方法学习并回忆配对词表，即要求被试用一个句子把每对词联结起来记；(3)用单个项目复述的方法学习并回忆 10 个项目的词表，即一个项目接着一个项目的重复；(4)用多个项目复述的方法学习并回忆 10 个项目的词表，即将多个项目结合为一组来复述。四次实验的词表都不相同，并使组内各实验的顺序平衡。学习与测验结束时要求被试回答：你选择哪一种方法来学习，为什么选择这一种方法而不选择别的方法？

第三阶段：自由选择方法的学习与回忆。在配对词表学习与回忆、非配对词表学习与回忆前，告诉被试他们可以使用在做游戏和学习单词时学到的方法。

① Lodico, M.C., Ghatala, E.S., et al. (1983). The effects of strategy monitoring training on children's selection of effective memory strategies. *Journal of Experimental Child Psychology*, 35, 263—277.

测验完成后问被试:你选择了哪一种方法来学习,为什么选择这一种方法而不选择别的方法?

实验的结果显示:(1)在各种实验条件下实验组被试的回忆成绩均更好。毫无疑问,这与实验组在实验前的游戏阶段进行的训练有关,这种训练是针对策略监测进行的,而且也强调了儿童要在学习中关注他的成绩。(2)实验组有更多的儿童把较好成绩归因于有效策略的使用(特别是对较难的非配对词表的学习任务),控制组没有儿童把较好成绩归因于有效策略的使用。(3)第三阶段,实验组与控制组被试在选择有效策略的百分数上存在显著差异。这说明训练对策略选择产生了明显影响。在要求这些儿童回答他们为什么选择这种策略时,实验组儿童大都回答说因为这种策略能够提高记忆成绩,而控制组儿童大部分不能说出自己的选择理由,或者给出的解释不正确。训练的效应达到非常显著的水平。

策略监测训练的核心是提供有关策略的有效性的认识和强化主动选择策略的意识,同时它也对形成积极的认知自信心有帮助,因为训练中可以使被训练者正确地把记忆的成败归因于认知策略的有效性。

董奇在1988年进行了一项有关儿童阅读的元认知训练实验。[①]该实验的被试是小学四年级和初中二年级的学生,而且实验与学校的日常教学相结合。培养措施主要包括两方面:一是在阅读教学中注意丰富学生关于阅读的元认知知识;二是注意培养学生对阅读过程的实际监控能力。实验采取的主要措施有:(1)让学生充分认识阅读的特点和实质;(2)让学生充分了解影响阅读效果的三个主要因素:个人因素、材料因素和策略因素;(3)在阅读时如何制定计划;(4)在阅读过程中如何进行有效的监控和调节;(5)阅读后如何检查结果、获取反馈信息、采取相应的行为。经过一年左右的训练实验,实验组学生关于阅读的元认知知识和元认知监控水平都有非常显著的提高,而且元认知训练促进了被试思维品质的发展。

张厚粲等人(1994)对LOGO程序设计教学在儿童元认知发展过程中的影响进行了为期一年的实验。[②]实验中,富有经验的教师在进行程序设计教学中有意训练学生的元认知能力。具体做法是:(1)教给学生如何制定计划、选择策略、

① 董奇:《中小学生阅读能力与元认知的发展与培养》,北京师范大学博士论文,1988。

② 张厚粲、吴正、宋华:《LOGO程序设计教学对儿童元认知能力发展影响的研究》,《心理发展与教育》,1994,1:1—5。

提出假设、检验假设并根据反馈对原计划进行调整等一系列有效解决问题的方法;(2)通过各种练习,训练儿童在解决程序设计问题的过程中,灵活运用上述方法,同时也训练儿童对解决问题的自我监控能力。这一教学促进了学生元认知能力的发展。

概括上述实验,我们看到比较有效的元认知训练,都必须先从培养和训练儿童使用具体的认知策略开始,并在训练中给学生提供充分练习的机会,让儿童充分了解各种认知策略的意义、原理和使用各种策略的相关知识,使有关认知策略知识成为儿童知识结构的有机部分。兰格和皮尔斯(Lange & Pierce, 1992)根据元认知的基本结构和功能,以及大量有关元认知训练的实验和试验,提出了一个精细的认知策略监控训练程序。①他们为促进儿童认知策略的习得与保持,将训练分成6个步骤:(1)提供一个关于策略的理性认识,即告知儿童一个策略为什么能对提高成绩产生积极影响;(2)提供以成绩为基础的反馈,使儿童确信这些策略的有效性;(3)提供足够的策略练习机会;(4)使用不同的刺激系列证实策略的普遍有效;(5)提供各种不同学习材料的策略练习机会;(6)有意激发儿童在作业中的参与意识。他们使用这一程序训练的学前儿童在“学习-回忆”中的策略行为都有明显增长。这一程序可以简单地表示成:讲解示范——试验反馈——练习强化——迁移训练——激发主体意识。②很明显,这一程序遵循了个体学习的一般过程和规律。

三、眼动心理学的实验研究

近年来,随着国内一批“眼动仪实验室”的建成,眼动心理学的实验研究成为国内实验心理学的一个新热点。目前,心理学中的眼动实验研究主要集中在阅读和图形认知方面,就教育心理学领域来说,研究最多的还是阅读的眼动特征。

(一) 眼动及其记录技术的发展

眼动(eye movement)即眼球的运动,它有三种基本方式:注视(fixation)、眼跳(saccade)和追随运动(pursuit movement)。③注视是指将眼睛的中央窝对准某

---

① Lange, G. & Pierce, S. (1922). Memory-strategy learning and maintenance in preschool children. *Developmental Psychology*, 453—462.

② 邓铸:《元记忆教育训练的认知研究》,《华东师范大学学报(教科版)》,1997, 4:68—73。

③ 阎国利:《眼动分析法在心理学研究中的应用》,天津教育出版社2004年版。

一物体的时间超过 100 ms,在此期间被注视的物体成像在中央窝上,获得更充分的加工而形成清晰的像。但注视不等于眼球的静止,注视中常常伴随着三种极为细微的眼动:自发性的高频眼球微颤、慢速漂移(slow shift)和微跳(micro-saccade)。这些细微眼动是视觉信息加工必需的信息提取机制。一般认为,慢速漂移使目标逐渐离开中央窝的中心,而由微小跳动纠正这个偏差,以保持正确的注视状态,这就使得被注视物体在视网膜上的成像位置不断发生变换以克服网膜适应导致的视像消失①,而眼球微颤则可将信息刺激调制成交流信号以便能穿过视觉通道。②

眼跳是注视点或注视方位的突然改变,这种改变个体往往意识不到。眼跳速度快,最高可达每秒 450 度,眼跳幅度可从 2 分到 20 度。眼跳过程帮助个体获取刺激的时空信息,但几乎不能形成刺激的清晰像,所以眼跳可以实现对视野的快速搜索和对刺激信息的选择。

追随运动是当被观察物体与眼睛存在相对运动时,为了保证眼睛总是注视这个物体,眼睛会追随物体移动。追随运动常常伴随较大的眼跳和微跳,它是由于运动目标的速度信息输入到中枢神经系统,眼睛为了追随这个目标而引起的一种连续反馈的伺服运动。

上述三种眼动方式经常交错在一起,目的均在于选择信息、将要注意的刺激物成像于中央窝区域,以形成清晰的像。眼动可以反映视觉信息的选择模式,对于揭示视觉信息的加工机制有重要意义。研究发现,两眼的眼跳几乎完全一致,这为眼动研究提供了诸多便利。那么,眼动记录的参数有哪些,如何得到这些眼动参数呢? 从近年来发表的研究报告看,利用眼动仪进行心理学研究常用的资料或参数主要包括以下几种。

1. 眼动轨迹图

眼动轨迹是将眼球运动信息叠加在视景(visual scene)图像上形成注视点及其移动的路线图,它能最具体、直观和全面地反映眼动的时空特征③,由此判定

---

① Riggs, L. A., Ratliff, F., Cornsweet, J. C., et al. (1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society American*, 37, 415—420.

② 刘伟、袁修干:《人的视觉—眼动系统的研究》,《人类工效学》,2000,(4):41—44。

③ Sperati, C. (2003). The inner working of dynamic visuospatial imagery as revealed by spontaneous eye movements. In J. Hyona, R. Radach & H. Deubel (Eds.), *The mind's eye: cognitive and applied of eye movement research*. Elsevier Science BV., 119—142.

不同刺激情境、不同任务条件、不同个体间、同一个体不同状态下的眼动模式及其差异性。

2. 眼动时间

将眼动信息与视景图像叠加后，利用分析软件提取多方面眼动时间数据，包括注视（或称注视停留）时间、眼跳时间、回视时间、眼跳潜伏期①、追随运动时间以及注视过程中的微动（包括自发性高频眼球微颤、慢速漂移和微跳）时间。同时，可以提取各种不同眼动的次数，主要是在不同视景位置或位置间的注视次数、眼跳次数、回视次数等。这些时间和位置信息可用于精细地分析各种不同的眼动模式，进而揭示各种不同刺激情境下的信息加工过程和加工模式。

3. 眼动的方向和距离

眼动的方向和距离就是在二维或三维空间内考察眼动方向（角度）。这方面的信息与视景叠加可以揭示注意的对象及其转移程度，而且可以结合时间因素计算眼动速度。结合眼跳角度计算注视广度②，$\mathrm{tg}\,\theta = D/S$（$\theta$ 代表眼跳角度，D 代表注视广度，S 代表眼球到阅读文本之间的距离）。注视广度反映出读者在一次注视中注视内容的多少，与注视次数大体呈负相关，是考察读者阅读能力的一个重要指标。

4. 瞳孔大小与眨眼

瞳孔（pupil）通过括瞳肌和缩瞳肌调节进入眼球内部的光线总量，并与其他视网膜及皮层细胞构成瞳孔反射（或称光反射），从而推动眼的折光成像。瞳孔直径的变化反映了读者认知加工时的心理负荷情况。贾斯特和卡彭特（Just & Carpenter, 1993）认为句子结构会影响瞳孔直径，即结构复杂时，瞳孔直径变大，反之瞳孔直径缩小。③瞳孔大小与眨眼也是注意状态的重要指标，而且与视景叠加可以解释不同条件下的知觉广度，也可以揭示不同刺激条件对注意状态的激活。

眼动的时空特征是视觉信息提取过程中的生理和行为表现，其与心理活动

① Huey, E. B. (1980). *The psychology and pedagory of reading*. The Macmilian Company, 9, 572.

② 沈德立：《学生汉语阅读过程的眼动研究》，教育科学出版社 2001 年版，第 171 页。

③ Just, M. A. Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indicates of sentence processing. *Canadian Journal of Experimental Psychology*, 47(2): 310—339.

直接或间接的关系奇妙而有趣，这也是100多年来心理学家致力于眼动研究的原因所在。眼动记录最初是作为生理学任务进行的。德国著名生理学家缪勒、休克(Hueck)等曾在19世纪对眼动进行了系统研究，发现了视旋转，即眼球以视轴为中心的中央旋转，揭示头部转向一侧的运动是补偿眼睛在相反方向上旋转的规律。后来沃尔克曼(Volkman)尝试对眼动的速度进行测量，验证了眼睛水平运动比垂直运动的速度快，单眼运动比双眼运动快。此后，眼动技术随着科学技术的发展而发展，先后经历观察记录法、机械记录法、光学记录法、影视记录法等几代方法。现在使用的主要是集成了光学技术、摄影技术、计算机硬件技术和计算机软件技术的眼动记录系统——现代化大型精密的眼动仪。国内也有系列的眼动仪产品，但其商品化进程缓慢。心理学研究中使用的还多为进口眼动仪，加拿大SR公司生产的EyeLink-Ⅱ眼动追踪记录系统、美国“应用科学实验室”(Applied Science Laboratories，简称ASL)研发的504和501型眼动仪、德国SMI公司研发的SMI眼动追踪系统(SMI Eye Tracking Systems)等，都是集成的眼动追踪系统。

现代眼动仪的结构一般包括四个系统：光学系统、瞳孔中心坐标提取系统、视景与瞳孔坐标叠加系统、图像与数据的记录与分析系统。①第一，光学系统。这是在一个特定的支架结构中装载三个微型摄像机，其中一个指向视景、一个对准左眼眼球或来自左眼眼球的反射光路、一个对准右眼眼球或来自右眼眼球的反射光路(有的只装载两个摄像机，除一台用于拍摄视景之外，就只有一台摄像机记录一只眼睛的运动)。像加拿大SR公司生产的EyeLink-Ⅱ是在头盔式支架上装载三个摄像机，其中有两个摄像机直接对准左右眼眼球；德国SMI公司生产的iView X Hi-Speed型眼动仪是在固定式支架上装载两个摄像机，其中一个对准来自左眼或右眼眼球的反射光路。第二，瞳孔中心坐标提取系统。该系统主要利用图像扫描及其信号转换，将眼动录像中瞳孔中心方位的信息转换成标准二维坐标的数字信号，然后传送给计算机处理并贮存。第三，视景与瞳孔坐标叠加系统。这是一个计算机的软件系统，它能将表征瞳孔中心及其移动的数字信号进行适当变换，与来自视景摄像机传送的数字信号匹配，将瞳孔中心坐标叠加在视景图像上，从而直观地标记出某一时刻瞳孔中心对准场景的哪一个位置(即注视点)，以及随着时间的推移注视

① 邓铸：《眼动心理学的理论、技术及应用研究》，《南京师大学报(社会科学版)》，2005，1：90—95。

点的移动路线。第四,图像与数据的记录与分析系统。这也是一个计算机软件系统,它能将叠加后的图像呈现在显示器上,并将其数字化以表征注视点的时空特征。由此,可以得到的主要结果和参数包括:注视点随时间变化的轨迹图,注视点分布的三维和二维密度图,各种不同形式眼动的位置、速度和时间等等。

(二) 中文阅读的眼动研究

国内关于中文阅读的眼动研究包括句子水平和语篇水平的研究。我们通过几项典型的实验研究来说明阅读的眼动研究的过程、数据记录与数据分析方法。

1. 句子阅读的眼动研究方法

"根据语言学概念,句子是表达完整思想的具有一定语法特征的最基本的言语单位。"①它以字词识别为基础,通过对字词的物质外壳进行编码,然后进行句法分析和语义分析,从而揭示句子意义。②但是由于句子加工的速度快、时间短,句子理解的实验研究长期存在技术制约,直至 20 世纪 60 年代以前,研究者很难采用实验方法对句子分析作定量研究。不过,从 20 世纪 60 年代开始,梅勒等人(Mehler, Bever & Carey, 1967)最早把眼动技术应用到阅读中的句法分析,随后积累的研究资料逐渐使人们相信"眼动资料能够反映读者在理解句子时对句法信息的即时加工过程"。③

句子阅读的眼动研究一般采用两种范式:一种范式是对结构不同的句子的阅读比较,另一种范式是对不同被试群对句子阅读的比较。研究中一般都是向被试呈现某一句子,让被试尽量用较少的时间来阅读,阅读结束后完成理解检测项目,从而揭示不同句式对阅读的影响,或者研究不同被试对相同句子阅读中存在的差异。实验的具体过程一般包括以下三步。

第一步:实验材料的制作与被试选取。根据研究目的,制作实验用语句,如歧义句、双重否定句等,并抽取适当的被试。沈德立(2001)等人研究中使用过包含有暂时结构歧义的句子:

a. 师长提拔了团长器重的士兵。

b. 连长提拔了团长器重的士兵。

① 高明凯、石安石:《语言学概论》,中华书局 1985 年版。
② 沈德立:《学生汉语阅读过程的眼动研究》,教育科学出版社 2001 年版,第 321 页。
③ 同上,第 324 页。

在动词“提拔”后的名词“团长”既可解释为宾语，又可解释为宾语的修饰成分，这个名词为结构歧义词。歧义词之后的“器重的士兵”是解歧区，它能消除歧义，表明“团长”一词只能作为宾语的修饰成分。上述a句的前半句语义合理，b句的前半句语义不合理，被试在对这些语句的阅读中有什么不同的视觉信息提取特征呢？这就构成了实验目的。

如果研究中想考察不同性质的被试在阅读句子中的差异性，则可以选取不同的被试来参加实验，收集他们的眼动信息并进行比较。

第二步：实验实施与眼动记录。阅读的眼动研究一般要用高速的、固定式的眼动仪来完成，如 iView X Hi-Speed 眼动追踪系统①、4200R 型眼动仪等。这些眼动仪的使用过程一般都是这样的：首先，让被试进入实验室，大致了解实验室的环境后，坐在眼动仪前的椅子上，向其介绍实验的大致过程和任务；接着，调整被试座椅的位置和高度，使被试能比较好地把下颚放在固定支架上，要求其注视眼前的显示器，调整眼球摄像头的位置和焦距，以及相应部位的状态。上述工作完成后，要对其注视点进行定标，即要求被试先后注视显示器上的若干点，以设定眼球注视点坐标记录的位置系统。调整完成后，要求被试不要移动头的位置，即可进入实验记录程序，在正式记录之前，可以做些阅读练习。这里需要注意实验指示语的记录和呈现，我们以沈德立(2001)等人实验的指示语为例来说明：②

这是一个关于阅读的实验，主要目的是要看你是怎样阅读句子的。一会儿我将在屏幕上呈现一个句子，请你按平时的阅读速度阅读，读完后立即报告“完”。在你读完一个句子后，接着屏幕上会呈现一个检查句，如果这个句子与你刚才读过的那个句子的意义相符，请你回答“是”，不相符则答“否”。在阅读句子时要默读，不要出声。在整个实验过程中，要保持头部不动。下面请你读一些句子(呈现供练习用的句子)，按我刚才提出的要求去做。明白实验要求了吗？如果有什么不明白的问题请提出来。如果没有问题，下面我们开始做实验。

① iView X Hi-Speed 眼动追踪系统为德国 SMI 公司生产，目前该设备的中国总代理为北京伊飒尔界面设计公司，该公司的网站地址为：http://isaruid.com。

② 沈德立：《学生汉语阅读过程的眼动研究》，教育科学出版社 2001 年版，第 333 页。

准备就绪后，开始呈现刺激语句，同时启动眼动记录程序，将被试阅读语句时的眼动信息记录下来。为了保证眼动记录的准确性，在每个句子呈现前要在屏幕上呈现一个信号（接下来呈现的句子是从此处开始的），并指示被试首先注视此位置。

第三步：眼动信息参数的提取与分析。实验结束后，利用眼动分析软件对眼动记录资料进行分析，这种分析一般从整句和局部两方面进行。整句分析中一般使用的指标包括：整句的阅读时间、注视次数、注视点持续时间、眼跳距离、注视频率和回视次数等。对句子阅读中一些局部眼动信息的分析，多采用"兴趣区"(area of interest，简称 AOI)方法，即依据句子结构将句子分解成小的单元。如上例中研究者将实验语句均划分成三个兴趣区。

表 10-2　句子阅读中兴趣区的划分举例

实验中使用的原始语句	兴趣区划分		
	兴趣区 1	兴趣区 2	兴趣区 3
师长提拔了团长器重的士兵	师长	提拔了团长	器重的士兵
连长提拔了团长器重的士兵	连长	提拔了团长	器重的士兵

有了具体的兴趣区后就可以将眼动信息数量化，提取各个兴趣区的一系列眼动数据：①

第一次注视的时间：指在该区域内从左向右的注视时间；

第二次注视的时间：指除第一次注视时间之外的对该区域的注视时间，包括回视时间和再次阅读该区域内容的时间；

兴趣区总的注视时间：指对该区第一次注视时间和第二次注视时间的总和；

注视次数：阅读中进入该区的次数。

有了上述数量化的结果，就可以采用适当的统计手段对其进行分析，包括对不同被试或被试组之间的眼动数据差异性分析、不同结构句子之间的眼动数据的差异性比较、不同实验控制或操纵条件下的眼动数据的差异性比较等，由此可以揭示句子阅读的特点、影响因素和内在心理机制。

2. 语篇阅读的眼动研究方法

国内关于语篇阅读的眼动研究虽然涉及多种文体，如叙述文、科普文、寓言

① 沈德立：《学生汉语阅读过程的眼动研究》，教育科学出版社 2001 年版，第 324 页。

故事等，但研究程序和从眼动记录中提取的信息或参数基本一致。①②③语篇眼动研究的程序一般如下。

首先将阅读材料制作成短文，短文长度最好能在电脑显示屏上一次完全显示，同时准备若干测试项目，以便在阅读后对被试的理解程度进行测试。

做好实验准备，包括编写好实验指示语之后，引导被试进入实验室，按照与句子阅读实验相似的操作程序进入实验过程，这一实验中一般需要两名主试，其中一人负责刺激材料呈现的控制，一人负责眼动记录操作。

实验中，当每一材料呈现后，被试进行阅读，眼动仪同时启动记录程序。阅读方式可以是加以控制的，也可以是不加控制的。控制性阅读就是对阅读顺序、阅读时间加以限制，如要求被试按文字顺序并出声阅读，限定被试在若干时间内读完等；不加控制的阅读，就是对被试的阅读顺序和阅读时间均不加限制，让其自己决定何时结束阅读。

被试读完一个材料后，主试呈现测试项目以检测其理解程度，然后进入下一个阅读材料的阅读、记录和实验过程。

实验结束后，研究者就可以利用眼动分析软件对实验记录进行分析，将眼动信息转换为一系列数据并使用统计软件进行分析。

语篇阅读中，常用的眼动参数主要包括以下几个方面。

(1) 阅读时间、阅读速度和阅读效率

阅读速度是指单位时间内的阅读字数，阅读效率是指在单位时间内的阅读理解率，其计算公式是：阅读效率$(E)=$阅读速度$(R)\times$阅读理解率(C)。阅读理解率是指阅读理解的正确率。

(2) 注视次数、注视点持续时间

人们在进行阅读时发生的眼球的连续运动被称为眼跳，两次眼跳之间眼球的相对静止状态被称为注视，一次注视也被称为一个注视点。注视次数是指注视点的数量，注视点持续时间是指每个注视点的平均注视停留时间。

(3) 眼跳和回视

眼跳的作用是为了改变注视点，使新刺激置于中央凹视觉区，以便看得清

---

① 白学军、沈德立：《初学阅读者和熟练阅读者阅读课文时眼动特征的比较研究》，《心理发展与教育》，1995，11(2)：6—12。

② 阎国利：《不同年级学生阅读科技文章的眼动过程研究》，《心理科学》，1999，3：225—228。

③ 陈向阳、沈德立：《中小学生阅读寓言过程的眼动研究》，《心理科学》，2004，27(4)：777—780。

楚。眼跳距离的大小在一定程度上可以反映一次注视的内容多少。回视(regression)是指被试眼睛的注视点从右向左的移动,但这不包括换行移动的情况。

(4) 信息区分析

信息区(area of information,也简称为 AOI)的划分。已有语篇阅读的研究,多以一个字或词为一个信息区来分析读者阅读过程,以便提取被试阅读每个字或词的时间和次数。这种划分信息区的方法可以了解读者对每一个字词加工的情况,但也存在一些问题,因为有研究发现人们在阅读文章时,有些词被跳读或快速掠过(Just & Carpenter, 1980),可这些跳读的词也被加工(Fisher & Shebilske, 1985)。如果我们仍然以一个词或字为信息区,那么就会低估了读者阅读的速度。

(5) 阅读方式

阅读方式的划分是按被试阅读时的眼睛注视顺序进行的。当把阅读材料和问题同时呈现给被试时,被试阅读时的眼动方式就会出现以下几种不同形式:(1)直接阅读方式,指被试从阅读材料的开始一直读到材料最后,然后再去考虑回答问题;(2)问题式阅读方式,指被试先注视问题,再注视阅读材料;(3)往复式阅读方式,指被试先从材料开始处阅读,中间不断地在行间回视或重复阅读过的内容,然后阅读问题;(4)跳跃式阅读方式,指被试在阅读过程中突然跳过某一句或数句。

(6) 瞳孔直径

瞳孔直径(pupil diameter)变化幅度。瞳孔直径变化主要反映被试认知加工的紧张度。

眼动研究是一个方兴未艾的领域,其技术手段、研究思想和涉及的课题领域都还处在迅速发展过程中。尤其是在中国,随着各大学纷纷开办心理学专业和投入大量资金建立包括“眼动仪实验室”在内的实验研究基地,了解和准备使用眼动仪进行心理学研究的心理学家队伍将会在近几年内迅速扩大。

阅读材料 10-1

现场研究的边界与结构特点

心理学的生命力在于应用,而传统的实验室方法很难准确、充分地反映实际

生活和生产活动中人的心理过程的许多边界条件和背景因素。因此,在心理学研究中,无论是教育、儿童和临床还是管理、社会和工程等心理学分支,现场研究日益成为主要方式。它不但进一步验证了实验室条件下的某些结果,而且更广泛深入地了解实际生活中众多的心理现象。

现场研究的背景十分不同于实验室条件,由于是在实际工作、学习和生活情境中进行研究,因而包含了各种复杂因素。现场背景的下列特点是心理学研究中必须认真考虑的:

1. 边界特点

(1) 变量强度

在现场背景中,某些重要的自变量可以显示出较高的强度,比如应激情境;而在实验室背景中,由于伦理和其他原因,某些自变量的强度受到限制。

(2) 变量变化范围

在现场背景中,许多变量的变化范围远远大于实验室条件,例如,群体规模、管理幅度、组织结构集中化程度、复杂性和时间广度等因素和变量的变化范围,都比实验室中更大。此外,跨文化现场研究的范围更是如此。

(3) 变量操纵的频次和持续时间

由于在实验室条件下研究时间往往比较短,人们常常有能力缓冲较短时间中的各种刺激,而且,实验室中不易提供包括各种反应阈限的条件。正因为如此,许多心理效应在实验室中没有被察觉,而在现场情境中却十分明显。某些小刺激会由于反复出现或持续较长时间而产生意外的心理效应。变量操纵的频次多和持续时间较长,这是现场研究的又一边界特征。

2. 结构特点

(1) 时间结构

许多心理过程都是在一定的自然时间广度上展开的,现场研究情境为我们提供了长时间结构和自然时间单元,而这些特征对于真正掌握心理规律又是十分重要的。

(2) 自然单元

这是指在现场条件下,研究者可以寻求特定环境下人们行为的自然单元,即自然发生的过程,这就避免了实验室条件所造成的人为背景与结构的影响。

(3) 复杂性

现场背景比实验室条件要复杂得多,许多情况在实验室里不可能模拟,因

此，现场背景提供和保持了研究复杂心理过程的条件。

以上三种结构特点，都是在实验室研究中为了控制条件而常常设法消除的，因而使研究的效度受到影响。

——资料来源：《现场研究的特点和条件》(王重鸣，1990)①

阅读材料 10-2

实验儿童心理学在中国的发展

在中国，由于历史的原因，20 世纪 70 年代末心理学才恢复。80 年代初，在一次全国性的儿童心理学学术会议上，笔者在一篇题为《关于儿童心理实验中控制因素的问题》的学术报告中，根据自己多年的研究实践，提出要建立与发展我国的实验儿童心理学。②

笔者认为，就全国范围来说，经过十多年来我国儿童心理学工作者的共同努力，目前已经具备了在中国建立与发展实验儿童心理学这一学科的各种基本条件。这些基本条件是：(1)目前国内已经有一支适合从事实验儿童心理学研究工作的专业队伍。(2)国内有一些单位(如北京大学、华东师范大学和天津师范大学等)可以研制与生产心理学实验所需要的仪器，为实验工作提供先进的技术手段。(3)社会发展的需要。1999 年，《中共中央、国务院关于全面推行素质教育的决定》明确提出，要改革旧的教育思想、教育内容和教学方法，必须重视学生心理素质的教育与培养。许多重要的教育问题急需心理学研究来提供理论依据，如大脑功能的开发与人才培养，创造力的培养，学习效率的提高等。这些都需要开展实验儿童心理学的研究。

研制心理学仪器是开展实验儿童心理学研究的基础条件。这是因为要保证研究成果的科学性，除了提高研究者的实验设计思想，制定出周密细致的研究方案之外，还必须在技术上解决两个问题：(1)能够精确地控制呈现刺激的时间和数量；(2)能够准确地记录儿童的各种反应。为此，笔者所在单位——天津师范

---

① 王重鸣：《心理学研究方法》，人民教育出版社 1990 年版，第 104—105 页。

② 沈德立：《关于儿童心理实验中控制因素的问题》，见全国高校儿童心理学研究汇编：《当前儿童心理学的进展》，北京师范大学出版社 1984 年版，第 161—173 页。

大学从20世纪80年代初开始，运用现代电子技术研制了第一代国产心理学仪器10种。这些仪器设计科学，各项指标符合心理学原理的要求，具有一机多能、创新性和实用性强的特点。20世纪90年代中后期，又研制了第二代国产心理学仪器——心理实验台。该实验台是对第一代国产心理学仪器的集约化和现代化，可以开展30余项心理学实验。目前，由北京大学、华东师范大学和天津师范大学等单位研制与生产的国产心理学仪器达50余种，供应全国高校和科研单位使用。

进入21世纪，中国心理学界面临了许多新的历史机遇。首先，心理学科被国家科技部列为国家在2001—2015年期间优先发展的18个学科之一。其次，第28届国际心理学大会将于2004年在北京召开。最后，心理学已经在社会生活中得到普遍的关注和认可。但是，对于实验儿童心理学研究来说，还必须解决好以下几个问题：

1. 实验儿童心理学的研究方向问题。没有正确的方向，就可能产生很大的盲目性，甚至带来不良的后果。什么是正确的方向呢？就是：(1)在理论上必须坚持辩证唯物主义和历史唯物主义思想；(2)在实践上必须坚持为我国全面建设小康社会和改革开放服务；(3)在吸收国外心理学资料上，必须贯彻“洋为中用”的方针；(4)在发掘、整理中国古代的心理学思想时，要坚持“古为今用”的原则。

2. 进一步探索心理发展规律的问题。科学研究的任务就是发现规律，这样才能为社会实践服务。心理发展的规律问题主要包括：(1)先天与后天的关系问题；(2)教育与心理发展的关系问题；(3)心理发展的动力问题；(4)年龄阶段特点与个别差异问题等等。

3. 实验儿童心理学研究的方法问题。在实验儿童心理学的研究方法上，必须做到以下三点：(1)坚持综合研究与分析研究相结合；(2)坚持专业队伍研究与群众队伍研究相结合；(3)坚持传统研究手段与现代技术手段相结合。这样才能促进实验儿童心理学的学术水平的提高和学术队伍的发展。

——资料来源：《实验儿童心理学》(沈德立，白雪军，2004)①

建议阅读文献

1. 黄希庭：《心理学实验指导》，人民教育出版社1988年版，第403—485页。

① 沈德立、白学军：《实验儿童心理学》，安徽教育出版社2004年版，第31—33页。

2. 郭力平:《学前儿童心理发展研究方法》,上海教育出版社 2002 年版,第 3—25、239—260 页。
3. 郑金洲:《学校教育研究方法》,教育科学出版社 2003 年版,第 102—142 页。

复习思考题

1. 如何理解成熟因素、被试流失、纵向研究、横断研究、临床实验法、元认知、注视、眼跳、追随运动、眼动仪、现场研究?
2. 发展心理学研究的特殊性有哪些?
3. 如何保证发展心理学研究的外部效度和内部效度?
4. 教育心理学研究的特殊性有哪些?
5. 对儿童心理进行实验研究时应注意哪些问题?
6. 分析比较纵向研究和横断研究的优缺点。
7. 举例说明元认知训练的一般步骤。
8. 眼动研究的常用参数有哪些?
9. 中文阅读眼动研究的一般步骤。

第十一章
人因工程学研究

本章内容提要

人因工程学与人机工效学、工程心理学、人类工效学等概念是近义的或有相当的交叉，是20世纪中期前后开始发展、80年代后开始繁荣的新兴学科，其在西方国家已有深入研究并在社会生活的各种工程设计中有广泛应用。本章在介绍其发展简史之后，对相关概念进行了比较分析，认为现代工业生产或工程设计必须是基于人因的考虑，取“人因工程学”最为贴切。在中国，大陆地区此方面研究尚处于繁荣前的时期，除少数几家研究机构外，“人因工程”概念尚不为公众所了解，相比之下，台湾地区的研究更为全面和具体。此外，本章还简要地介绍了一些人因工程学研究方法，主要涉及人体尺寸参数测量方法、生理状态与环境检测方法、作业疲劳检测方法和仿真实验技术等。

国内公众对“人类工效学”(ergonomics)有一些了解，但对“人因工程学”(human factors engineering; ergonomics)就很感陌生，以至不少新华书店把《人因工程学》类的书籍摆在了生物学、基因工程类的书架上。从词源看，二者几乎相同。当前文献中出现的“ergonomics”更多地被理解或翻译为“人类工效学”，但有国内学者开始将其翻译为“人因工程”或“人因工程学”。这一翻译上的改变，虽不能动摇迄今为止国内该领域研究目标的核心——追求“工效”，但实际上已经反映了该领域的系统性变化：“人的因素”越来越成为“人-机-环境”研究中关注的重点，同时该领域研究越来越强调“人-机-环境”系统中人的安全、健康和舒适。在美国，“人的因素”成为与“人类工效学”相平行的一项研究，其重点明显放在实验心理学和实验工程学方面。“ergonomics”一词来自希腊文，其中的“nomo”是“准则”之意，“ergo”是“工作”之意，组合起来就是指有“准则”的“工作”，为此，“ergonomics”应该提供具有前瞻性的设计。设计是工程范

畴，而要使设计具有前瞻性，就必须以“人的因素”为基础。因此，我们倾向于使用“人因工程学”的概念，这一概念也有了自己更为直接的英文术语“human-factors engineering”。

第一节　人因工程学概述

一、人因工程学的发展简史

(一) 西方国家的人因工程学

泰勒(Frederick Winslow Taylor, 1856—1915)被认为是最早对人与工具匹配问题进行研究的学者。19世纪末，他在美国的伯利恒钢铁公司进行了一系列提高工作效率的试验，找到了一种能帮助工人最有效地铲运煤、铁矿石的铁铲型式。公司采用这种铁铲后，工人的劳动效率提高了好几倍。随后，泰勒开创的时间研究与吉尔布雷思(F. B. Gilbreth, 1868—1924)开创的动作研究被合称为动作时间研究。[①]不过，这一时期该领域更多关注的是人员选拔与培训问题，真正属于人因工程范畴的研究还很少。

我们知道，第二次世界大战前的一个世纪，是西方国家资本积累的重要时期，这一时期的劳工问题异常突出。一些矿道、工场的工作条件和工作时数的安排很少考虑劳工的安全与健康，国际社会和一些国家的政府开始制订并通过一系列法案以保护劳工，特别是1919年“国际劳工组织”(International Labour Organization，简称ILO)成立后，这些法案开始对劳工工作场所进行规范，这可被认为是人因工程实践的开始。

如果以第二次世界大战为界，人因工程学的研究可被划分为两个历史阶段。战前为第一阶段，该阶段主要解决人对机器的适应问题。闵斯特伯格(Hugo Münsterberg, 1863—1916)等心理学家用心理学的方法和原理为已经生产出来的机器选拔和训练操作人员，选拔与训练使操作者适应已成型机器的性能。特别是第一次世界大战期间，交战各国都征召心理学家为军队选拔飞行员、潜水员等特种兵员。

随着科学技术的进步，机器不断更新换代，其结构与性能日益复杂，而人的

① 朱祖祥、葛列众、张智君:《工程心理学》,人民教育出版社2000年版,第13页。

素质提高却总是有限度的。所以,完全通过人员选拔和训练去适应不断发展的高性能机器终究是行不通的。第二次世界大战中,各交战国竞相发展新的高性能武器装备,例如飞机的飞行速度有很大提高,飞行高度向超低和超高扩展,武器火力系统加强,在复杂气象条件下执行任务的机会增多,信息显示器和控制器的数量急剧增加。面对如此迅速发展的技术水准,即使经过最严格训练的飞行员也难以胜任,如此一来,不仅不能充分发挥先进武器装备的性能,而且更频繁地发生着诸如机毁人亡的严重事故。研究人员逐渐意识到,工具与机器的设计不是一个孤立事件,它必须考虑到使用者的能力限度,必须考虑到与人的特点最好的匹配,才能充分发挥其性能。于是,欧美参战各国开始把人因工程学运用在武器设计发展上,研究如何使武器的使用更为有效、安全,保养更为简便。所以说,真正有系统地对人因工程进行研究,开始于第二次世界大战。

第二次世界大战结束后,工业发展的最大目标就是发展生产力。而这也是人因工程学的一个很大目标,因为工业生产力和员工的表现有直接关系。例如,装配零件的速度、搬运物品的速度等都会影响工作绩效。渐渐地,机械力开始取代人力,机器操作成为工人的主要工作方式。而且机器的力量越来越巨大、运转速度越来越高,这同时意味着,错误的操作所带来的伤害和后果更为严重。所以,战后美、苏、德等国的一些大公司陆续建立了人因工程学研究机构。20 世纪 60 年代,工业的发展和人因工程的目标也由发展生产力逐渐转移到提升机器安全性上,工厂的生产方式也由小组生产演变为以一个流程的方式生产。工人的角色开始由直接动手工作转变为监测或检查流程。在这种情况下,发生意外的频率降低,因为工人离机器执行动作的地方较以前远多了,但是,意外一旦发生,后果却更为严重。人因工程师在生产线的设计中解决一个又一个问题,使其越来越适合人的特征与身心限度。

此外,人因工程还在产品设计中发挥着作用,使产品设计越来越适应或符合使用者的需求与特点。就汽车工业来说,人因工程对他们的重要性不只是在生产线,还包括对驾驶、乘客、维修人员所做的考虑。现在的标准程序里都包括不断对产品做人因工程的测定,如在汽车工业中就包含对行车、座椅舒适度、操控性、噪声、振动程度、操作的简易性、视野大小等各项因素的考虑。

20 世纪 70 年代,国外人因工程学处于快速发展阶段,诸如居住、工作环境、消费品的设计,医疗,保健,娱乐以及生产过程,品质监管都列入研究范围,逐渐呈现繁荣态势。而且,这一时期的研究比以往更加关注人类自身的因素,综合提

升人类生活品质、工作舒适度、安全度，以此为前提来提高生产力，正是人因工程学应有之方向。

（二）我国人因工程学的研究

我国工业界和学术界开展的人因工程学研究，主要在“工程心理学”、“人类工效学”的麾下进行，近年才开始以“人因工程”的名称来表述。[①][②]20 世纪 30 年代，清华大学开设了工业心理学课程。《工业心理学概观》(陈立，1935)是我国最早系统介绍工业心理学的著作。该书作者当时还在北京南口机车车辆厂和无锡大通纱厂等处开展工作环境和择工等问题的研究。新中国成立后，从 50 年代中期开始，中国科学院心理研究所和杭州大学等单位的心理学家在职工培训、操作合理化、预防工伤事故等方面做了许多工作。这一时期的工作在方法上多采用调查、观察、现场试验等。到了 60 年代，我国工业生产中的各种装备由仿造为主发展到自行设计制造为主，这就要求心理学提供人机匹配的数据，从而促进了我国工业心理学的发展。一批工程心理学家先后在仪表设计、铁路信号显示、电站中央控制室、航空照明和显示等方面开展了研究。遗憾的是，在 1966—1976 年，由于众所周知的原因，研究工作停顿，直到 20 世纪 70 年代后期，我国工程心理学的研究才开始恢复和发展。中国科学院心理研究所、航天医学工程研究所、空军医学研究所、杭州大学、同济大学等单位先后建立了工效学或工程心理学研究机构。杭州大学还创建了工业心理学专业，招收和培养了第一批工程心理学的本科生、硕士生和博士生。一些高等学校开设了工程心理学或工效学课程。国家技术监督局和国防科工委分别组建了“全国人类工效学标准化技术委员会”和“军用人-机-环境系统工程标准化技术委员会”。有关学科的专家为制定人类工效学标准开展了多方面的人-机-环境关系问题研究，取得了大量研究成果。

目前，中国大陆地区，人因工程研究在工业安全与卫生、工业设计与人机界面两方面的情况比较好。前者主要是因为中国大陆地区正处在工商业急速发展的阶段，工场安全问题日益突出，相关的护具、装备、标志等产品从国外进口过于昂贵，此方面的开发空间都还比较大。人机界面则与信息软硬件、家电产品密切相关，因为市场规模庞大，国内厂商开始在开发产品上投注资本，自然而然面临

① 郭伏、杨学涵：《人因工程学》，东北大学出版社 2001 年版。
② 蔡启明、余臻、庄长远：《人因工程》，科学出版社 2005 年版。

着工业设计与人机界面设计的需要;国际品牌为了本土化也不得不投注研究资源。本土厂商多为转民营的国营企业或单位,以及改革开放之后崛起的民营企业,如联想、海尔、TCL、康佳等,这些企业已经迅速累积雄厚资本,成为中国的名牌,还进一步朝国际品牌迈进,这些中国名牌相当重视产品的开发。市场需要逐渐激活人因工程在国内的研究,但是迄今为止,此方面的研究还多是以市场开发为导向,学术性研究相对较少。

近几年,随着市场经济的快速发展,一些严重的工场安全事件也频频发生,这引起政府部门的高度重视,加强了对工场安全问题的研究与监管。同时,政府部门还加强了对相关领域或行业的工效学国家标准的制订与颁布,成立了相关的监管机构,如隶属国务院的"国家安全生产监督局"。国家标准则由标准化研究院组成的"全国人类工效学标准化技术委员会"进行制订,委员会邀请产官学界代表数十人参加,目标是依据 ISO 系列的国际标准建立与国际接轨的中国国家标准,包括人类工效学一般指导原则、人体测量与生物力学、人-系统交互工效学、物理环境工效学、照明工效学以及劳动安全工效学等。

在学术研究方面,浙江大学心理与行为科学学院沈模卫教授领导的研究群体、中国科学院心理研究所张侃研究员领导的研究群体进行了较长期的研究,取得了丰富的研究成果,成果多发表在《心理学报》、《人类工效学》、《应用心理学》等相关学术刊物。此外,清华大学工业工程学系、北京航空航天大学、北京大学公共卫生学院也都在开始相关领域的研究。

需要强调的是,在中国台湾地区,人因工程学的研究已经形成相当规模,有包括台湾科技大学在内的多家大学建有专门的人因工程实验室,并有较为丰富的科学研究成果,我们将在下文做专门介绍。

二、人因工程学及相关概念的分析

该领域研究也在形成一个概念系统或概念群,如"工程心理学"、"工业心理学"、"人类工效学"、"工效学"和"人因工程学"、"人因学"等。如果简单归纳一下,这六个概念明显属于三组同义词:前两个术语限定了较为狭窄的研究范围,强调了"工程"与"工业",多取"工程心理学";中间两个术语强调了生产线或工具产品的"操作工效",多取"人类工效学";最后两个术语强调了工程设计中"人的因素",多取"人因工程学"。下面,就对这三个概念进行比较。

(一) 工程心理学的界定

朱祖祥教授新近给“工程心理学”下的定义是:“工程心理学是一门心理学与工程技术交叉的学科。它从心理学的观点研究工程技术、制造和使用中人与技术的关系问题。”①在进一步的论述中,他认为工程心理学就是为了研究如何使工程技术产品与使用者的身心行为特点相适应的问题而发展起来的科学。这样的学科,在美国曾经被叫做“应用实验心理学”(applied experimental psychology),20 世纪 50 年代中期又被称为“人类工程学”、“人素工程学”。目前,还有不同学科介入的研究者,对该学科的称谓也不一样,如“人机工程学”、“人-机-环境系统工程学”等。

工程心理学的研究目的主要在于三方面:(1)研究人、机器和环境三者之间的关系,以此为基础研究如何设计出能与人满意匹配的机器、工具,以提高工具使用的效率,提高投资效益;(2)研究机器操作中人的因素,尽可能避免事故发生中的“人为错误”,提高工场的安全性(简称“工安”),防止事故和保障安全,特别是操作人员的生命安全。历史上有过许多惨痛的教训,多数都是“人为错误”造成的,如 1986 年发生在乌克兰北部的切尔诺贝利核电厂 4 号核子反应炉爆炸事故,经历 4 年的调查后,将当时因被认为负有错误操作责任的有关人员释放,因为这些操作人员本来就不具有足够的相关知识,在故障处理中出错是无法避免的。有研究者认为其中的关键问题在于:操作程序的设计者没有很好地认识到操作人员的知识水平,所以现代工程设计中必须考虑到操作者的知识水平,并使系统具有较高的“容错性”;(3)创造健康和舒适的工作条件,这需要研究与人的身心健康、舒适感有关的因素,如劳动强度、负荷、噪声、照明、振动、温度和湿度等。

(二) 人类工效学的界定

人类工效学是应用科学领域中的一门学科,研究个人受到他所使用的工具、机器及其他装备、具体环境的影响,借以帮助人们更容易、更有效、更舒适和更安全地进行工作。“人类工效学”这一术语创始于 1949 年,用以表示人体解剖学、生理学和实验心理学与工程学的物理学相交叉的一个领域。人类工效学是一个较大的概念,它还包括一系列子学科,如管理工效学、人机工效学、环境与安全工效学、认知工效学、交通工效学、工效学标准化以及人类工效学应用等方面。

① 朱祖祥、葛列众、张智君:《工程心理学》,人民教育出版社 2000 年版,第 1—4 页。

与工程心理学相比，人类工效学涵盖的范围更大一些，它超出了工程设计领域，但核心还在工程设计，即在工程设计中追求更好的“工效”。

（三）人因工程学的界定

人因工程学是一门将心理学、生理学、解剖学、管理学以及各种工程学知识与方法相结合，探讨人、工作和生活彼此关系的新领域，它的主要功能在于研究如何设计一个最适合人的生活和工作的新环境。欧美各国，大都将人因工程学归于工业工程研究的范畴，经过几十年的发展，它已经可以在工业化以及公众生活品质的提高方面发挥重要作用。

从本质上说，人因工程学的研究目标在于“以物就人”和“如何以物就人”。随着科学技术的快速发展和经济条件的快速改善，公众对各种商品的审美需求、对商品使用的舒适度的要求越来越明显，为此必须做专门的研究，即人因工程学的研究。比如，目前全球化的办公自动化发展很快，以亿计的工作人员每天要在计算机的终端机前工作，终端机在办公室的重要性及需求，将仅次于电话。以往，人们对终端机的研究，注意焦点集中在辐射对人体的影响方面。事实上，尚有许多重要因素被忽视，而这些因素往往对办公效率产生很大影响，如键盘、屏幕大小、角度、座椅和工作台的高度，以及舒适的程度、屏幕显示的颜色、光度和字体大小、分辨率等。简言之，一个良好的办公环境必须考虑以下因素：

（1）座椅的高度、靠背、脚垫、手垫是否合适？

（2）终端机的屏幕角度、光度、对比、颜色是否合乎人类视觉要求（以视觉不易疲倦及减少错误为标准）？

（3）键盘的角度、字键排列顺序、距离、大小、表面摩擦力的大小都必须考虑；

（4）所有设计都要以在一定范围内能调整为原则，以适应个体差异；

（5）外在环境，如温度、湿度、空气流动的情形、噪声、工作压力也必须注意。

以上要考虑的设计必须以人因为基础，着眼点在于配合人的体型设计出最舒适的办公环境，尤其是自动化的办公场所，而这就需要对人类体型进行测量和研究。就目前来看，中国人因工程学的研究资料并不丰富，有许多工程设计依赖外国人的人因资料，但外国人在体型方面又往往与我们中国人很不相同。不难推断，我们正在使用的许多产品和工具不适合中国人的人因的程度是严重的，中国急需建立自己的人因资料体系，也就是要建立自己的人因工程学研究体系。

很明显，上述三个概念几乎是同一的，因为它们在研究对象、研究目的、研究方法和手段、研究的内容方面基本一致，只不过强调的重点稍有不同。这种强调重点方面的细微差异又随着研究者越来越多的共识而更加不明显，不同研究者使用着不同的术语，这在相当程度上是因个人的习惯不同而已。如果我们强调，现代工业生产或工程设计必须基于人因的考虑，那当然取“人因工程学”（human factors engineering）最为贴切。

第二节　国内的人因工程研究机构

中国的人因工程研究尚处于发展初期，未形成规模，即使是在局部地区、个别领域特别受到重视①，但在公众中的影响度还很小。心理学科中，也只有中国科学院心理研究所、浙江大学心理与行为科学学院两家单位有专门的研究队伍，并在培养硕士生和博士生。此外，台湾地区则建有包括台湾科技大学在内的多家“人因工程实验室”，比大陆地区的专门研究机构要多，研究也更系统和更接近市场。为形成对人因工程学研究的感性认识，本节介绍中国有代表性的研究机构和实验室。

一、浙江大学的“工业心理学实验室”

原杭州大学心理学系是我国工程心理学的最重要研究基地，该基地是在我国心理学大师陈立（1902—2004）教授领导下建立起来的，以后又涌现出朱祖祥等著名心理学家。后来杭州大学并入浙江大学，心理学系也改称为“浙江大学心理与行为科学学院”，学院建有国家级“工程心理学实验室”，由沈模卫教授主持工作。“工业心理学实验室”是国家教委于1989年批准在工业心理学重点学科点的基础上建立的，1990年通过国家教委和国家计委的评审，开始接受世界银行“重点学科发展项目”的设备和技术援助贷款，至今已有实验用房2 000平方米、固定资产600万元，包括“视觉追踪”、“临境技术”、“航空工程心理学”、“人-

① 毫无疑问，在航天领域，我国的人因工程研究已经达到了相当高的水准，已经在实践中实现了两名宇航员的安全进入太空、安全飞行一百多个小时、安全返回地面，而且宇航员“感觉良好”。不过，这一特殊领域的许多研究成果是不便公开发表的。

计算机界面"等一批追踪国际先进水平的实验室。主要设备包括:眼动仪、1980B光谱辐射测试仪、PR-1530微光测试仪、精密多功能声级计、SGI工作站局域网等先进的硬件设施以及MultiGen和Vega等国际一流的三维建模和仿真软件。实验室承担了多项国家级研究项目,包括"863"项目、攀登计划项目、国防科工委重点科研项目等。研究领域或内容主要包括航空工程心理学、人机界面设计、人体测量和工作负荷研究等。

(一) 航空工程心理学研究

目前承担国防科工委、航空工业总公司及其下属研究所委托的多项研究课题,主要开展飞机座舱电子信息综合显示与控制的工效学研究、座舱总体设计的工效学研究、视听综合告警研究、座舱眩光预防研究,以及其他技术设计和使用中人的因素问题研究。

(二) 人机界面领域的研究

主要承担过"863"项目"幻境技术与系统研究"和国家自然科学基金重点项目"用户界面评估和人机对话系统及虚拟座舱实例"等研究项目,主要开展人机对话系统和多通道用户界面设计评估的研究、虚拟现实系统研究。其中"863"项目的研究成果"虚拟故宫"的开发,追踪了国际上先进的虚拟现实技术,为国内的虚拟现实技术的发展奠定了基础。

(三) 工程人体测量和工作负荷研究

工程人体测量方面,主要研究工程设计人体测量方法,以及为座舱等设计提供所需人体功能尺寸;在工作负荷方面,主要研究工作负荷评价方法,工作负荷与操作可靠性的关系,以及特定人机环境系统中人的工作负荷评价等。

(四) 认知心理学等基础理论方面的研究

主要研究人的信息加工能力和在高技术条件下的人机关系模型,以推动心理学理论的发展。目前主要承担汉字字型识别的信息加工研究、中文新闻记者过程的心理学研究、视错觉的研究等多项国家和省自然科学基金项目。

二、中国科学院心理研究所

中国科学院心理研究所的荆其诚、林仲贤、陈永明等心理学家都曾进行了一系列基础性、应用性研究,其中不少内容涉及人因工程问题。目前,张侃研究员

领导的研究工作基本上都属于人因工程领域①,主要包括工程心理学②、人与计算机的交互作用③④、人因学(人类工效学,ergonomics)、人的空间认知、注意的机制、人的智力工作负荷等。

三、清华大学工业工程系“人因工程实验室”

人因工程实验室主要用于培养学生人机交互方面的有关工程与心理知识和实际运用技能,该实验室主要包括生理基本参数测量与作业设计、心理测试与认知特性分析以及有关复杂作业过程虚拟再现三方面内容。人因工程实验室分为四个区域,分别为基础实验区、人体测量与生物力学教学实验区、虚拟现实与人机界面教学实验区和可用性评测与人机交互教学实验区。人因工程基础实验室主要针对本科生基础人因工程课程教学,满足基础实验教学要求。主要实验内容包括人体基本测量(形体尺寸、生理参数、认知特性等)、环境基本测量(温度、湿度、光照、噪声、辐射、空气等)以及工作地基本测量(几何、物理测量)等,使学生掌握基本的人因学测量手段和测量概念,强化人因工程的概念,并且能在今后的实际工作中自觉地考虑人-机-环境的互动关系,提高工程设计的人性化水平。

虚拟现实与人机界面技术实验室主要针对以虚拟现实技术为基础的复杂人-机-环境交互系统进行模拟、仿真、测评和研究,立足于构建各种人机界面进行近似真实场景的测评,并运用多通道人机界面技术开展部分虚拟体验教学研究,通过视觉、听觉、触觉的集成体验来加深理论知识的学习并感知复杂的系统或理论。本实验作为高等人因工程专题实验和一系列研究类专题实验的重要部分,培养学生运用新技术解决复杂问题的思路和基本能力,特别是自己动手进行研

① 张侃,1990年在美国伊利诺斯州立大学获心理学博士学位,现任中国科学院心理所研究员、博士生导师;中国心理学会理事长、国际心理科学联合会(IUPsyS)执委、国际应用心理学会(IAAP)执委、国际人因学会(IEA)执委、国际科学理事会(ICSU)科学自由常务委员会委员、国际科学理事会(ICSU)中国委员会副主席和中国科学院心理研究所学术委员会主任。

② 张侃:《刺激—反应相容性原理与工程心理学研究》,见中国心理学会编:《当代中国心理学》,人民教育出版社2001年版,第513—519页。

③ Zhao, C. ,Zhang, T. & Zhang, K. (2001). User interface design for the small screen display. *Proceedings of the Human Factors and Ergonomics Society 45th annual meeting*, pp. 1548—1550.

④ 吴昌旭、张侃:《诊查型用户界面可用性评价方法(IM)(上)—简介与评价》,《人类工效学》,2000(3)。

究类实验的设计、运行、总结、提高的能力。同时该实验室还作为工业工程系其他专业方向教学和研究用实验平台，开展体验类教学实验和成果展示。主要内容包括三通道立体虚拟现实投影系统、触觉人机交互装置研制、虚拟现实软件平台购置以及专题实验设计。

四、上海交通大学"系统可靠性与人因工程实验室"

该实验室以船舶控制室为实验背景，提供实时的信息采集，能完成 1 000 多个实验数据点的仿真、测试，提供船舶模拟实验环境、光照等，并配备了相应的计算机设备，有简单的灯光控制系统以及照度测量设备。

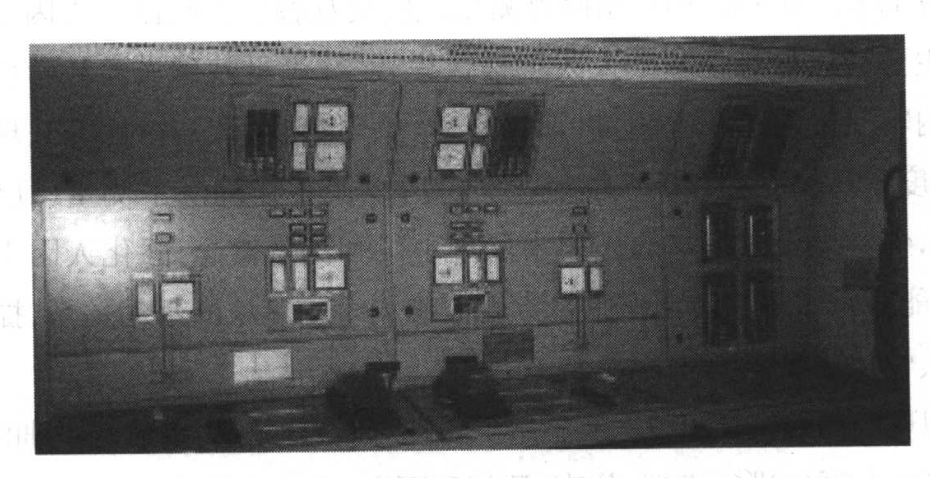


图 11-1　上海交通大学"系统可靠性与人因工程实验室"

可完成船舶舱室及海洋平台控制室模拟实验，进行人因工程分析研究，对提高人的操作效率及设计优化进行评价。图 11-1 所示的就是为实验室船舶机电集控室仿真装置。

实验室中安装的控制面板仿真装置，可完成人机界面操作及船舶仿真控制显示。实验室已纳入上海交通大学船舶学院"211 工程"计划，在 1999—2000 年建成由模块化可变换机架、模块化可变换面板、多功能中心控制计算机组成的综合控制模型。其中机架是设备的主体，可以通过改变模块来实现不同形式的组合，包括高低、显控位置等；可变换面板将分别包含两个显示器以及各种仪表、控制手柄、指示灯等常用的显控媒体及部件；多功能中心控制计算机则负责各种显示信号的发送以及人员操作信号的收集处理。

实验室中还装备了多功能模块化人因工程试验台，这是进行控制及驾驶系

统人因工程量化分析研究的有效工具。它通过计算机控制模拟被控系统运行，同时自动记录人的操作结果并进行统计处理，得到在一定条件下人的操作可靠性数据。可以对人因工程进行定量化评价。

该实验室研究取得的一些成果除直接用于船舶设计改进外，也发表了一系列学术论文。

五、台湾地区的“人因工程”实验室

在面临日益高度自动化、复杂化的操作系统下，作业人员心智负荷日益繁重，常导致人为疏失，影响系统安全。因此，人因工程致力于研究作业中的认知过程与影响作业绩效的因素，并将之应用于系统设计，以达到系统的高效化、可靠性和安全性。台湾地区工程学界较为充分地认识到了这一点，至少在以下大学建立了人因工程学实验室：台湾清华大学、台湾科技大学、台湾交通大学、中原大学、朝阳科技大学、华梵大学等。

(一) 台湾交通大学“人因工程实验室”

台湾交通大学的“人因工程学实验室”划分四个分室，包括计算机系统人机接口研究室(Human-Computer Interaction Laboratory)、认知历程研究室(Cognitive Processes Research Laboratory)、环境控制实验室(Environmental Factors Research Laboratory)、产品设计与人体工效学研究室(Product Design/Ergonomics Laboratory)。实验室的主要装备除常规的计算机、扫描和照相器材外，还有以下用于研究的主要设备：生物力学实验仪器(如行为分析器)、职业性向测验器、工作性格测量器、疲劳测定器、温度感觉测定器、团体反应分析器、尘埃测定器、时间规划器、刺激产生器、噪声计、速示器、混色仪、触觉器、亮度计、数据收集系统、腿力和臂力评估系统与曲度计等。

实验室研究的重点包括：大型或复杂系统操作中“人为错误”的研究、电子信息产品人机接口的设计、以使用者为中心的产品研发以及软件工程的辅助工具。具体地说，研究主要涉及以下五个领域。

(1) 人类认知与操作行为基础研究：对人的复杂认知行为如学习、思考、推理、创造及决策的本质与过程并进行模型建构，并探讨影响训练及学习效果的因素。包括教育模式的建立、作业环境的设计、团队合作、人际互动方式、反馈设计及不同设计效果的比较等。

(2) 过程控制系统研究:人为疏失肇事原因分析与经验反馈。台湾电力公司为了减少人为疏失所引起的跳机事故,与该实验室合作,应用有关人因知识,研发以防范人为疏失为目的的肇事原因分析工具,叫做"人员绩效增进系统"(Human Performance Enhance System,简称 HPES),研究制订了 TPC-HPES 技术手册,以供各核能电厂施用。

(3) 先进型主控制室人机接口研发:传统的核能控制室中那些令人眼花缭乱的仪表常使监控人员负荷过重,进步的人机接口设计策略即是以计算机呈现出整合的信息,以降低操作员撷取与整合信息时所需的心智负荷,因此,探讨不同信息整合显示形态对监控行为的影响也是研究的重点。

(4) 3C 产品的人机接口设计:3C 产品就是指计算机(Computer)、通讯设备(Communication)、商务电子产品(Consumer Electronics)。该实验室受相关政府部门的委托,进行台湾地区电子信息产品使用者的生活形态研究,此研究的重点在于由使用者的认知与使用行为,探讨各个电子信息产品间在使用上的关联性,并将其作为产品策略规划的基础和预测未来市场及使用者生活形态趋势的依据。此外,亦着手研究开发 Video-on-demand、家庭购物等使用者界面,以及遥控器的使用行为研究。

(5) 航空飞行系统研究:在航空飞行系统研究方面,主要包括对影响飞行安全问题的探讨、航管自动化的研究,飞行安全问题研究主要是研究飞航系统中出现问题的"人为因素",然后通过系统改良来避免各种"人为因素"导致的航空灾难。在飞航管制方面,为提升航管系统设备的自动化与现代化,应用人因方面的知识进行航管自动化的人机接口研究。该实验室拥有波音 747—400 程序训练仿真器,可以供仿真研究之用。

(二) 台湾清华大学"人因工程实验室"

台湾清华大学"人因工程实验室"分为四个分室:(1)工作生理实验室:配备"3D Scanner"、"EMG(肌电仪)"、"测力板"、"仪器升降台"、"三度空间量测仪"等仪器;(2)人体测量实验室:配备"三度动作分析仪"、"人体计测仪"、"人体姿势固定仪"等仪器;(3)作业环境实验室:配备"热压计"、"多用气体侦测器警报器"、"功率放大器"、"电子卡尺"、"耗氧量测试设备"、"心跳侦测表"、"噪声振动位准记录器"、"微波测定器"、"人体活动角度器"、"气体校正器"、"血压计"、"握力计"、"噪声振动器"等;(4)认知心理实验室:配备"视力检查仪"、"混色器"、"视觉测量仪"、"照度计"、"人体测量仪"、"视觉对比敏感度测试机"等仪器。该实验室

为学生主要开设了以下人因专业课程。

(1) 人因工程学:介绍人因工程的基本概念,了解人因工程的生理基础,工作场所设计,作业环境对人的工作绩效之影响,并在教学中研讨人因工程有关的原则与程序、如何设计适合人类使用的设备与工作环境等基本问题。

(2) 工业安全与设施规划:了解不同作业场所中不安全的因素及相关法令,进而学习促进现场工作安全的方法、技术以及管理策略,以提升整体系统的安全性;学习设施规划的基本理论、观念及方法,培养学生实际的设施规划能力。

(3) 人的信息处理与人机系统设计:了解各种人机系统模式的结构、适用范围及局限性,并能应用模式于系统设计;对人与计算机系统之互动作专题介绍,目的在于能使学生运用人因工程的专业知识进行自动化系统的通讯与工作设计;了解人的信息处理过程及其有限性等。

(4) 应用生物力学:训练学生进行人体测量、生物力学及工作生物学,以了解人的生理能力及限制,并能据此理解工作设计的基本思路。

(5) 工业环境分析:介绍有关现场作业环境安全性评估与设计方面的问题。其内容包括噪声、振动、照明、辐射、温湿等方面的测量与控制,以及个人防护器具的使用等。

(7) 人因工程测试与评估:了解人因工程中测试与评估的重要性及可用的方法,以确保所设计的人机系统能达成预定的目标。

该实验室的研究内容可从其近年承担的科研项目中得到反映:网络消费者的行为模式及应用在购物辅助界面设计的研究;3D数字人体模型的属性建构与仿真分析;无尘室生产环境下的人因危害评估与改善;符合人体工学螺丝起子的开发研究;建立轿车车门组装线数字化动作数据库与时间基表及动作编辑技术的应用;建构一辅助团队作业的计算机化模式;制造环境中人为失误损失推估模式的研究;LCD人类视觉计测与数据库建立;人体解剖标记点的自动标示法;人因系统接口效能评估技术的研究。从这些研究课题可以看出,实验室集中于人体参数测量、人机界面设计、工作场所设计与监测,以及灾难事故中“人为因素”的分析。

(三) 台湾科技大学“人因工程实验室”

台湾科技大学除直接命名的“人因工程实验室”外,“人机界面实验室”实际上也是人因工程研究的实验机构。

台湾科技大学"人因工程实验室"由李永辉教授主持①,故又称"李永辉人因工程实验室"。实验室装备主要包括:EMG 肌电仪②、ECG 量测仪③、QUARK PFT 肺功能测定仪、脚踏机、脑波仪(或叫脑电仪)、足底测力板、自动抬举机、雷射测距仪、道路实验车等。利用这些设备,实验室开展了一系列的产品设计④、现场改善、作业负荷等相关研究⑤。

"人机界面实验室"由纪佳芬教授主持,研究内容主要涉及重大职业灾害或事故原因分析研究、人机界面设计研究和残障人员机具及其交互界面的研究等⑥,如实验室装备了眼动监视系统、视觉仪、水晶体调节力测试仪、视野计、噪声计、照度计、疲劳闪值测定器、温湿度计等仪器。

从上述中国大陆和台湾地区实验室的介绍中,我们可以概括出国内人因工程学研究的主要方面,以及实验室设备配备的基本模式。概括地说,人因工程研究的主要方面:人体测量与环境监测及工作场所的改良;人-机-环境相互作用模式,包括界面设计的工效学问题;工作分析与人体生理心理能力的有限性问题;复杂人机系统事故中"人为因素"的分析与系统改良;安全、健康和舒适的工作场所的设计问题。为了对这些问题进行研究,人因工程实验室的常用设备包括四类:人体测量类、环境监测类、仿真实验类、心理与行为记录类。

第三节 人因工程学的基本实验方法

人因工程学研究的基本方法除前述各章介绍的实验方法,以及一般人文学科都采用的观察法、量表法之外,还包括一些较为特殊的方法和技术,主要有:人体参数与环境参数测量方法、仿真实验技术、人机系统事故中的"人为错误"分析等。这里简单介绍人体尺寸测量方法、人体生理参数测量方法、环境参数测量方

① 李永辉,男,美国得克萨斯州立理工大学工业工程博士,现任台湾科技大学教授并主持"人因工程实验室"的工作,台湾人因工程学会理事长。主要研究领域:工作改善、人因工程、风险管理等。

② 通过体表附接电极片,经由计算机联机收集肌肉活动信息的装置。

③ 测量被试在不同作业下的心脏负荷(心电图)。

④ 李永辉:《步枪重量操作长度对射击绩效影响之探讨,小客车第三刹车灯之装设位置对驾驶者视觉追踪行为之影响》,《工业工程学刊》,Vol. 10, No. 4, 251—256。

⑤ Lee, Yung-Hui. Effects of rifle weight, stock length and hand grip angle on aiming stability. *Applied Ergonomics*, 28, 2, 121—127.

⑥ 纪佳芬:《身心障碍者职务再设计与工作改善》,五南书局 2003 年版。

法和仿真技术。

一、人体体型参数测量方法

(一) 测量的内容及其人因学意义

人体测量学是一门古老而新兴的学科。说它古老,是因为有记载的人体测量工作开始很早,如在我国医学典籍《内经・灵枢》的《骨度篇》中已有人体测量的记载和阐述;公元前1世纪,罗马建筑师维特鲁威(Vitruian)为希腊神庙建筑研究了人体各部分的比例。意大利文艺复兴的伟大先驱达・芬奇(Leonardo da Vinci, 1452—1519)根据维特鲁威的描述绘制出了著名的人体比例图。[①]不过,古代和近代的人体测量,着眼点主要在于建筑、雕塑、绘画等。现代人体测量学则开始于19世纪末和20世纪初,最典型的是1919年,美国进行了一项对10万退伍军人的多项人体测量工作,并将测量得到的数据用于军服的设计与制作,这是真正意义上将人体测量应用于工业设计的开始。第二次世界大战后,美、英两国又进行了大规模的海军、空军人体测量,为大型战斗机具的设计提供了翔实的人体参数。

当前,各国为了设计的需要,都制订了本国的人体尺寸国家标准。我国也于1988年颁布了相应的国家标准——GB/T1000-1988《中国成年人人体尺寸》。

现代人体测量学主要是指通过测量人体各部位尺寸来确定个体间和群体间在人体尺寸上的差异,用以研究人的体态特征,从而为各种工业设计和工程设计提供人体尺寸参数。

人因工程范围内的人体尺寸测量包括两大类,即人体构造尺寸和功能尺寸的测量。人体构造尺寸是指静态尺寸;人体功能上的尺寸是指动态尺寸,包括人处在各种工作姿势或某种操作活动状态下测量的尺寸。

不难理解,各种器具和设施等设计对象在适合人的使用方面,首先要考虑的问题就是如何适合人的形态和功能范围的限度。例如,一切操作装置都应设置在操作者的肢体活动所能触及的范围内,其高低位置也要与人体相应部位的高度相适应;而且其空间布局应尽可能考虑人的操作方便、反应灵活和快捷,其目的就在于提高设计对象的宜人性,让使用者能够安全、健康、舒适地工

① 阮宝湘、邵祥华:《工业设计人机工程》,机械工业出版社2005年版,第32页。

作，从而有利于减少人体疲劳和提高人机系统的效率。在设计中所有涉及人体尺度参数的确定都需要应用大量人体构造和功能尺寸的测量数据。在设计时若不很好地考虑这些人体参数，就很可能造成操作上的困难和不能充分发挥人机系统效率。比如，就机器操作平台来说，其操作部位的高度与人的上肢舒适操作的高度相比过高或过低，人在操作时需要弯腰或抬臂，就会影响工作效率，也会造成人体的过早疲劳，长期操作还会对操作者的身体健康带来不利影响。①总之，人体测量参数对各种与人体尺度有关的设计对象具有重要意义。

当然，人体测量并非人们想象的那样简单，它必须在被测者姿势、测量基准面、测量方向、测点等符合一系列国际标准的前提下才算有效。在国标 GB3975-1983 中规定了工效学使用的有关人体参数的测点 22 个、测量项目 69 项，其中分为立姿 40 项、坐姿 22 项、手和足部 6 项，以及体重 1 项。在国标 GB5703-1985② 中又规定了工效学使用的人体参数测量方法，这些方法适用于成年人和青少年的人体参数测量，该标准对上述各测量项目的具体测量方法和各个测量项目所使用的测量仪器都作了详细的规定和说明。

（二）常用人体尺寸

GB/T10000-1988 是我国重要的人机工程技术标准，其中包括一系列人体尺寸数据标准，而这其中最常被引用的有人体主要尺寸、立姿人体尺寸、坐姿人体尺寸、人体水平尺寸、人体头部尺寸等五类。③

1. 人体主要尺寸

人体主要尺寸包括身高、体重、上臂长、前臂长、大腿长、小腿长。日常生活中，人们多关注身高和体重，但在人因工程学中，除这两项外，臂长和腿长与人工作、劳动的情况具有特别密切的关系，所以它与工程设计中如何提高工作效率、劳动效率有直接关系。

2. 立姿和坐姿人体尺寸

立姿人体尺寸主要包括眼高、肩高、肘高、手功能高、会阴高、胫骨点高六项。坐姿人体尺寸主要包括坐高、坐姿颈椎点高、坐姿眼高、坐姿肩高、坐姿肘高、坐

---

① 蔡启明、余臻、庄长远：《人因工程》，科学出版社 2005 年版，第 206—207 页。

② GB5703-1985，是指 1985 年颁布的国家标准（简称 GB），其编号为 5703。

③ 阮宝湘、邵祥华：《工业设计人机工程》，机械工业出版社 2005 年版，第 41—46 页。

姿大腿厚度、坐姿膝高、小腿加足高、坐深、臂膝距、坐姿下肢长。

立姿和坐姿条件下的人体尺寸很多，而国标中专门提供了上述尺寸数据，这是因为这些项目的尺寸对工作、劳动和器物设计而言相对更为重要。为理解这一点，我们摘引有关部分人体尺寸项目的应用说明(阮宝湘、邵祥华，2005)，如表11-1所示。

表 11-1　国标 GB/T10000-1988 中部分人体尺寸的应用

人体尺寸项目	应用举例与说明
2.1　立姿眼高	立姿下需要视线通过或需要隔断视线的场合，如病房、监护室、值班岗亭门上玻璃的高度，一般屏风及开敞式大办公室隔板的高度等，商品陈列橱窗、展台展板及广告布置等
2.3　立姿肘高	立姿下，上臂下垂、前臂大体举平时，手的高度略低于肘高，这是立姿下手操作工作的最适宜高度，因此设计中非常重要，轮船驾驶、机床操作、教师讲台高度等都要考虑
2.4　立姿手功能高	这是立姿下不需要弯腰的最低操作件高度；手提包、手提箱不拖到地面上等要求，均与这一人体尺寸有关
2.5　立姿会阴高	草坪的防护栏杆是否容易跨越、男性公厕中小便接斗的高度、自行车车座与脚踏的距离等，都与这一尺寸有关
3.1　坐高	双层床、客轮双层铺、火车卧铺的设计，复式跃层住宅的空间利用等与此尺寸有关
3.3　坐姿眼高	坐姿下需要视线通过或需要隔断视线的场合，影剧院、阶梯教室的坡度设计，汽车驾驶的视野分析，需要避免视觉干扰的窗户高度，计算机、电视机的放置高度，其他坐着观察的对象的合理摆放等
3.5　坐姿肘高	座椅扶手高度设计，与坐姿工作、坐姿操作有关的各种机器与器物，例如坐姿操作生产线工作台的高度，书桌、餐桌的高度设计等
3.6　坐姿大腿厚度	椅面之上、桌面抽屉之下的空间，是否容得下大腿或允许大腿有一些活动余地
3.8　小腿加足高	很重要，座椅椅面高度设计的依据
3.9　坐深	座椅、沙发坐深设计的依据

(摘自：阮宝湘、邵祥华，2005：43—44)

3. 人体水平尺寸

人体水平尺寸主要包括胸宽、胸厚、肩宽、最大肩宽、臀宽、坐姿臀宽、坐姿两

肘间宽、胸围、腰围和臀围等。

4. 人体头部尺寸

人体头部尺寸主要包括头全高、头矢状弧、头冠状弧、头最大宽、头最大长、头围、形态面长。

另外，需要说明的是，人体尺寸存在着很大的人种差异、民族差异、地区差异、性别差异和年龄差异，并且随着时代的变迁其变异很大，所以人体测量标准体系要考虑这一系列差异因素，建立系统的尺寸数据体系，服务日常生活和工程设计。

(三) 人体尺寸测量的仪器与方法

从技术发展看，人体测量技术可以分为普通测量技术和三维数字化人体测量技术。①普通人体测量技术就是采用一般的人体生理测量仪器，包括人体测高仪、直脚规、弯脚规、皮尺、测齿规、立方定颅器、平行定点仪等，其数字处理采用人工处理或人工输入计算机进行处理。这种数据处理费时费力、容易出错，数据应用不灵活。但是这种测量成本较低，在精度要求较低的情况下使用。

在要求比较精密的领域，关于人体参数的测试结果也要达到相当的精确度，而且数据采集量要大，这时就要采取先进的三维数字测量技术。三维数字化测量技术的基本原理，就是在人体表面选取一系列测量点，然后利用相关设备提取这些点的三维坐标。如果在人体表面的选点密度达到一定水平，就可以将这些点的数字坐标输入计算机，经过一定的软件处理，形成被测者体型的三维图像，从这些坐标就可以提取被测者的身高、臂长、体廓等许多外形特征。常用的设备有“三维数字化测量仪”、三维人体扫描仪等。如美国斯坦福大学设计生产的 Cyberware 全身 3D 扫描仪，能够依靠多只探头和计算机软件系统，在几秒至十几秒的时间内将人体三维体型信息提取出来并输入计算机，再由计算机进行数字处理得到人体体型特征参数。下面介绍的是三种通用的常规人体测量器具。

1. 人体测高仪

该仪器主要是用来测量身高、坐高、立姿和坐姿的眼高，以及伸手向上所及的高度等立姿和坐姿的人体各部位高度尺寸。

① 罗仕鉴、朱上上、孙守迁：《人体测量技术的现状与发展趋势》，《人类工效学》，2002(6)：31—34。

国标 GB5704. 1-1985 是人体测高仪的技术标准，该测高仪适用于读数值为 1 mm；测量范围为 0～1 996 mm 人体高度尺寸的测量。标准中规定的人体测高仪由直尺 1、固定尺座 2、活动尺座 3、弯尺 4、主尺杆 5 和底层 6 组成，如图 11-2 所示。

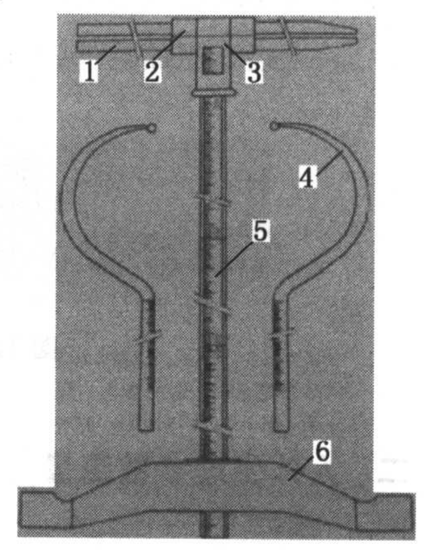


图 11-2　人体测高仪

1　直尺　2　固定尺座
3　活动尺座　4　弯尺
5　主尺杆　6　底层

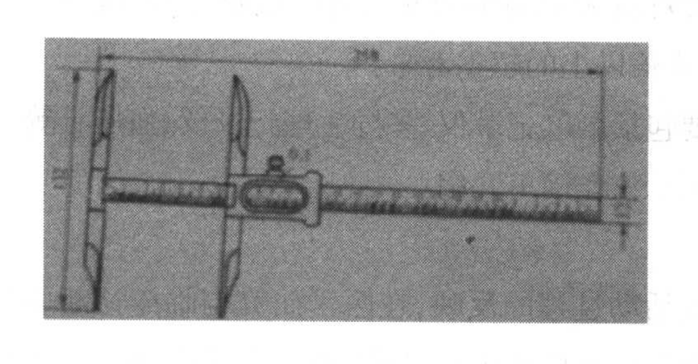
若将两支弯尺分别插入固定尺座和活动尺座，与构成主尺杆的第一、二节金属管配合使用时，即构成圆杆弯脚规，可测量人体各种宽度和厚度。

2. 人体测量用直脚规和弯脚规

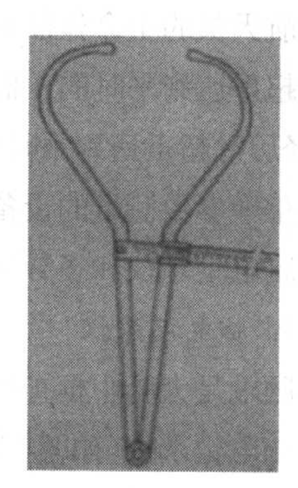
人体测量用的直脚规如图 11-3 中的 a 图所示，它由两个卡爪(前端的卡爪固定，另一个可以沿主尺移动)、主尺和活动尺座组成。直脚规是用来测量两点间的直线距离，特别适宜测量距离较短的不规则部位的宽度或直径，如测量耳、脸、手、足等部位的尺寸。

国标 GB5704. 2-1985 是人体测量用直脚规的技术标准，此种直脚规适用于读数值为 1 mm 和 0. 1 mm，测量范围为 0～220 mm 和 0～250 mm 人体尺寸的测量。直脚规根据有无游标读数分为Ⅰ型和Ⅱ型两种，其中有游标的精确度高一些。

弯脚规如图 11-3 中的 b 图所示，它由两个弯爪、固定尺座和一个主尺组成，主要用于不能直接以直尺测量的两点间距离的测量，如测量肩宽、胸厚等部位的尺寸。



(a) 直脚规



(b) 弯脚规

图 11-3　人体测量用直脚规和弯脚规

国标GB5704.3-1985是人体测量用弯脚规的技术标准,此种弯脚规适用于读数值为1 mm,测量范围为0～300 mm的人体尺寸的测量。

除上述常规器具外,人体测量中常用的还有角度计和软尺,如图11-4所示。

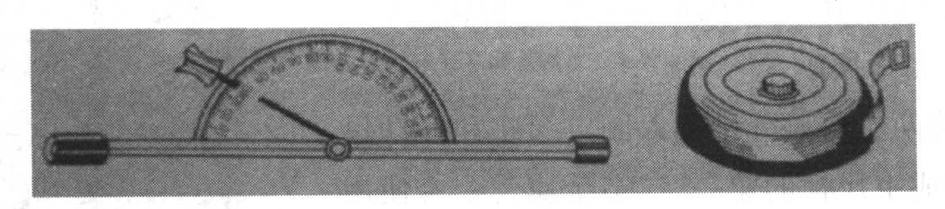


图11-4 人体测量用角度计和软尺

二、人体生理参数测量

体型测量为人因工程设计提供了基本的物理尺度,但是在人机相互作用与相互依赖过程中,人的身心状态的变化也是非常重要的,而心理变化有时又可以通过生理指标来反映,所以人因工程研究中的一项基本任务就是生理参数测量。主要包括血压及其变化范围、脉搏率及其变化范围、肺活量、脑电变化、皮肤电变化、抗撞击性等。这些方面的测量设备已经发展到相当的精度,而且可以实现遥测。比如,中国"神舟五号"和"神舟六号"载人飞行过程中,地面控制中心可以对航天员的生命体征进行实时测量,以评估航天员的身心状态,以及飞行器的内环境是否对航天员的生命活动带来不利影响。还有,在飞船发射升空过程中,人体的超重情况以及伴随的生理变化参数。

研究表明,火箭发射过程中,航天员的超重达到平时体重的4倍,在这种情况下航天员的生命体征参数如何,都是航天人因工程研究中非常重要的测量任务。超重达到平时的4倍不会对航天员的生命安全造成威胁,因为他们在训练中,经历过超重达到平时体重4倍以上的超重考验。

生理参数测量的设备主要包括脑电记录仪、多导生理记录仪、血压计等。下面对脑电记录仪和多导生理记录仪作简单介绍。

1. 脑电记录仪

脑电是大脑中枢神经系统活动过程的反映,在医学研究与临床诊断中常常使用脑电记录仪(如图11-5)来记录大脑中枢神经系统的电活动过程。脑电活动具有不同的形式,主要包括自发电位和诱发电位。自发电位(Spontaneous EEG)是指在没有特定刺激的情况下,大脑神经系统活动自发产生的电位及其变化;诱

发电位(Evoked Potential, EP),或称诱发反应,是指神经系统在接受内外环境刺激信息所产生的特定神经电活动。

临床上,通常将诱发电位划分为外源性诱发电位和内源性诱发电位。外源性诱发电位主要是指由于外界刺激引发的电位变化,如视觉诱发电位、听觉诱发电位、躯体感觉诱发电位和运动诱发电位。内源性诱发电位则是由于个体积极参与某项活动,特别是参与认知活动的情况下获得的诱发电位,因此也称为事件相关电位(Event Related Potential,简称为 ERP)。脑电记录仪就是记录这些自发电位和诱发电位变化的仪器,它因此也成为当代认知神经科学研究中的主要设备。

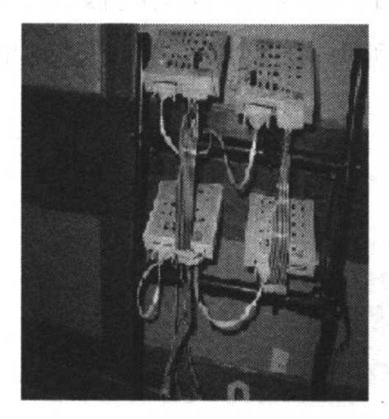
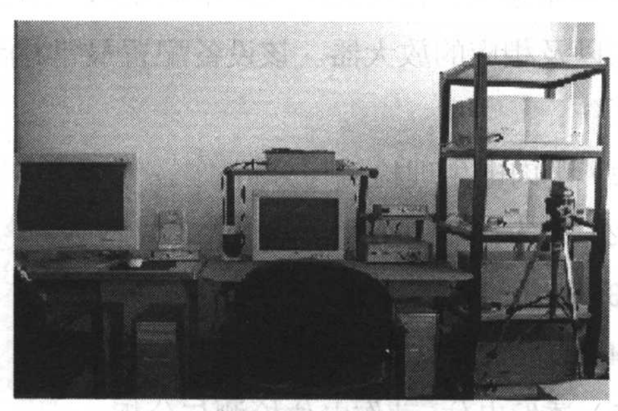



图 11-5　南京师范大学"认知神经科学实验室"配置的脑电记录仪

脑电记录仪的基本构成单元包括电极帽、刺激呈现器和被试任务操作平台、信号采集与放大装置、信号分析系统(由硬件和软件两大部分组成)等。

脑电记录仪的工作原理是:给被测试者戴上电极帽,安排其参与某项活动或完成某项工作任务,同时通过电极帽记录受测者头皮表面不同部位的电位变化,由此推断不同脑区的电活动情况,并进而推断其不同脑区的功能状态。

不过,目前脑电记录仪的空间分辨率还很不理想,其空间定位的精确性不够高。

2. 多导生理记录仪

多导生理记录仪是用来记录生理活动多项指标的医学与心理学实验仪器。根据多导生理记录仪的发展和研究者对不同生理指标需求情况的配置,它分为 4 导、8 导、16 导、32 导、64 导生理记录仪。研究者可以根据需要选用不同的配置。图 11-6 所示的是南京师范大学"认知与情绪实验室"配置的 16 导生理记录仪。

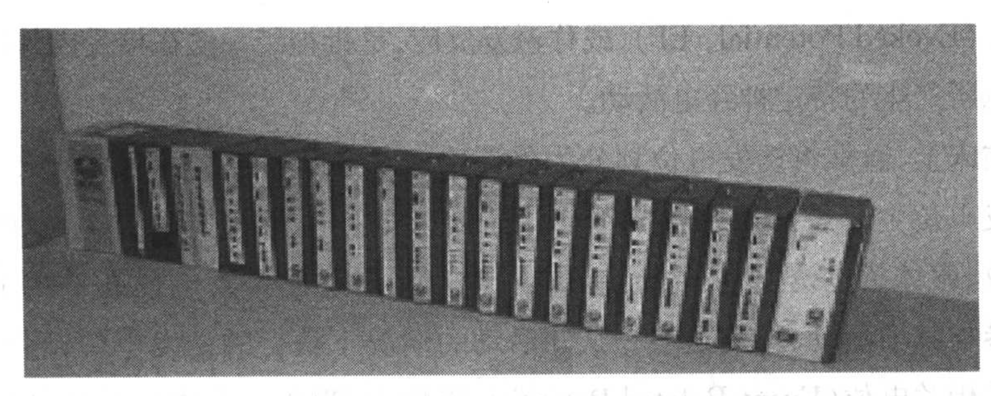


图 11-6 南京师范大学"认知与情绪实验室"配置的 16 导生理记录仪(BIOPAC)

多导生理记录仪主要由各种传感器、放大器、监视器和记录器组成。目前配备的传感器有用于记录肌肉生物电的电极、心电图电极、脑电图电极、流体压力传感器以及相应的放大器。该设备配置最常用到的,是采集并分析下列生理指标的系统:

(1) 肌肉电生理检测与分析;

(2) 人体心电图检测与分析;

(3) 脑电检测与分析;

(4) 皮肤电阻、皮温的记录与分析;

(5) 诱发电位、动作电位检测与分析;

(6) 呼吸、血压等张力压力信号检测与分析;

(7) 语音信号检测与分析;

(8) 神经元活动检测与分析。

由于生物学、医学技术的快速发展,人体生理参数的测量手段已经相当完备,在不同情境中可以根据需要选用不同的设备与手段。

多种常规生理参数测试的方法和应用价值已为公众所认知,而一些非常规的肌电、脑电测量也会在有关生理心理学或认知神经科学书籍中加以介绍,此处不赘述。但是,疲劳测量却是实验心理学或其他心理学课程都很少涉及,它又是人因工程研究中非常重要的一个方面,所以下面我们就疲劳的测量方法作专门介绍。

三、作业疲劳测量方法

"在劳动过程中,当作业能力出现明显下降时称为作业疲劳(fatigue),它是

机体的正常生理反应，起预防机体过劳(overstrain)的警告作用。"①人在处于疲劳时，常常出现从轻微的疲倦到精疲力竭的感觉，但是这种感觉并非总会伴随疲劳出现，比如具有高度责任感的人有时会"不知疲倦"地工作。疲劳会造成工作能力下降、应激能力下降，通常情况下会表现为作业成绩下降，如果仅此而已，倒也无妨。可是在某些特殊情况下，特别是在大型人机系统中，疲劳会带来严重后果，比如媒体经常报道因疲劳驾驶而出现的重大交通事故。这就使得疲劳测试成为非常重要的人因学测量任务。

疲劳有不同的类型或等级。②从类型方面看，疲劳包括：个别器官疲劳，如眼疲劳；全身性疲劳；智力疲劳，为长时间的脑力劳动导致的头昏眼花、乏力、失眠和易激怒等；技术性疲劳，常见于需要脑力、体力并重的，且需要注意力高度集中的工作，如驾驶汽车、飞机等。从疲劳感或疲劳发展来看，疲劳有三个阶段：轻微疲倦感、作业能力明显下降、强烈疲倦感。在一些常规的工作情境中，对生理状态的要求不严格，可以通过劳动者的自身感觉或观察其作业绩效就可对疲劳作出粗略判断。但在一些大型人机系统和复杂工作中，对生理状态要求很高，即使没有明显疲倦感的疲劳也会带来事故概率大幅增加，这时，我们就需要借助更严格的方法对疲劳程度进行测量或监测。

许多研究者认为，疲劳表现常有三方面的特征：(1)身体的生理状态发生特殊变化，例如，心率(脉搏率)、血压(压差)、呼吸及血液的乳酸含量等发生变化；(2)进行特定作业时的作业能力下降，如对特定信号的反应速度、正确率、感受性等能力下降；(3)疲倦的自我体验。

检测疲劳的方法主要包括以下三类：

1. 生化法

这种方法是通过检测作业者的血、尿、汗及唾液等成分的变化判断疲劳。这种方法的不足之处是测定时需要被测者中止作业，还容易带来被测者的不适和反感。

2. 生理心理测试法

除脑电、肌电、心率、血压等生理测定法外，常用的心理实验测定法主要包括：膝腱反射机能检查法、两点阈检查法、闪光融合临界频率测定法、连续色名呼

① 蔡启明、余臻、庄长远：《人因工程》，科学出版社2005年版，第67页。
② 同上，第67—73页。

叫检测法和反应时间测定法。其中,反应时间测定法已有专章介绍,这里也不再讨论。

(1) 膝腱反射法。使用医用小橡胶垂,按照规定的冲击力敲击被试的膝部,根据小腿弹起角度大小评价疲劳程度。被试的疲劳程度不同,引起反射运动的角度也不同,疲劳程度越大,反射角度越小。一般认为,作业前后反射角度变化5°~10°时为轻度疲劳,作业前后反射角度变化10°~15°时为中度疲劳,作业前后反射角度变化15°~30°时为重度疲劳。这些膝腱反射角度变化值也称为膝腱反射阈限。

(2) 闪光融合临界频率检查法。在被试工作前、后测试其闪光融合临界频率(CFF),并比较其变化。人体疲劳后,CFF降低,说明视觉神经出现钝化。这一方法对视觉显示终端前面的工作人员的疲劳测定最为适用。

(3) 两点阈测定法。两点阈就是指在用两个距离很近的针状物同时刺激皮肤表面两点时,被试能感觉到为刺激两点的最小距离。作业时机体出现疲劳时,皮肤感觉机能钝化,两点阈会增大。

(4) 连续色名呼叫检测法。通过检查作业者识别各种颜色,并能正确地呼叫出各种颜色色名的能力,判别作业者的疲劳程度。测试者准备5种颜色板若干块,相当快地抽取色板,同时让作业者回答。作业者若处于疲劳状态,回答速度较慢,且错误率相对增高。根据作业者的回答速度和错误率,可以判断作业者的疲劳程度。

四、环境参数监测

环境参数监测是评估人机系统以及工场条件的重要方面,它通过对环境各种参数的测量,来评估环境的安全性、健康性,主要就是对其中工作人员的安全性与健康性进行测量与评估。

通常的环境参数监测包括环境的温度、湿度、含氧量、有毒气体或物质分子的成分及含量,有无易致爆物质。比如,在煤矿的矿井内,由于煤层被掘开后会释放出多种有毒气体分子,因此要保证矿井的通风效果,避免有毒或易致爆气体的含量过高。经常发生的矿井瓦斯爆炸,就是由于没有做好矿井的环境参数检测与预报,其中有许多事故都是由于矿井瓦斯含量太高造成的瓦斯爆炸。

还有,像家居装修中的甲醛超标问题,已经成为居民健康的主要危害因子之一,也是当前社会生活中急需解决的人因学问题之一。

五、人机系统的仿真实验[①]

仿真实验也叫模拟研究，就是通过仿真或模拟技术创造出与研究问题的实际情形相同或相似的情境，使身处其中的被试与处于实际系统时的心理状态与行为活动相同或相似。一般来讲，仿真实验可以适用于各种人机系统研究和一些特殊工场设计的研究。当然，对于大型的人机系统来说，人机相互作用的程度比较大、相互依赖性比较强，而且造成的环境与日常人们的生活环境有较大差异，甚至根本不同，很难根据人们的日常表现去推定其实际处于复杂系统中的状态。因此，在没有积累起充分的研究资料或没有充分的把握之前，不能从一开始就将工作人员置于系统之中。最典型的就是太空飞行、深海潜艇作业和超高空飞行以及飞行特技表演等。

（一）仿真实验研究的意义

在人因工程学的人机关系仿真实验研究中，一般要用物理模型来代替机器或现场情境。这种物理模型系统有简单的，也有复杂的。简单的如模拟机车司机观察铁路颜色灯信号、模拟飞行员在某种照明条件下判读仪表。复杂的人机实验物理模型常常被称为模拟器或模拟室，例如飞机驾驶模拟器、太空飞行模拟舱、核电站中心控制模拟室等。这些模拟装置可以在相当程度上模拟出逼真的机器运行情境，工作人员身处其中可以完成各种与真实操作一致的动作，因此会产生出较为逼真的心理感受。

比如，飞行员在全景飞行模拟器中练习驾驶技能时，接受的信息和进行的操作，要求与飞机在天空飞行时发生的情形非常相似。初学飞行的人在实际驾驶飞机前，都要在模拟器上进行长时间的训练。模拟器的情境越接近实际飞行条件，飞行训练的效果就越好。在人机系统建设中，一个新的系统正式建造前，一般要先制造出模拟系统，研究人在模拟系统中的身心、行为反应特点，而后根据模拟试验取得的数据和发现的问题对系统设计作进一步的改进，再对改进的系统进行模拟试验，如此反复进行，直到被认为合理、安全、舒适为止，最起码要被认为是安全的。

我国“神舟五号”载人航天飞行之前的四次未载人发射也在某种程度上相当

① 朱祖祥、葛列众、张智君：《工程心理学》，人民教育出版社2000年版，第51—53页。

于仿真试验或模拟试验。在这些试验中，除了检验系统中的硬件、软件和通讯系统功能特性之外，还要检验其内部的物理环境是否适合生命存在，比如飞行舱中的温度、湿度、气体成分等。为了取得更为可靠的实验数据，在未载人试验飞行的同时，让受训航天员进入地面模拟舱，完成各种模拟操作和接受生理、心理测试。这些仿真实验为飞船的改进提供了大量资料，为保证"神舟五号"和"神舟六号"飞行过程中飞行员的安全起到了非常重要的作用。

仿真实验的重要特点是它具有现场研究的真实感，但不受现场条件的限制。在条件控制上，模拟实验与实验室实验的要求完全相同。它可以根据研究的需要控制实验条件，可以根据研究者的愿望设置和调整实验变量，参加实验的被试也可以根据需要加以选配。所以，仿真实验既克服了研究情境的失真问题，也克服了准实验条件控制的不严密性，是人因工程学领域中最重要、最常用和最有效的研究方法。

（二）人机系统仿真实验的逼真度

用物理模拟研究人机关系问题，其效果在很大程度上取决于模拟装置的逼真度。逼真度越高，得到的结果越接近实际情形。所以，单因素模拟不如多因素模拟，静态模拟不如动态模拟。

人机关系仿真实验的逼真度包括设备逼真度、环境逼真度和心理逼真度三个方面。设备逼真度是指模拟物的外观特点、信息显示器、操作手键的位置与操作方式同真实装备的相似程度。环境逼真度指模拟实验中的内外物理环境因素与被模拟的实际环境的相似程度。心理逼真度是指实验中被试在心理上认为模拟过程与真实情形的相似程度，它在很大程度上取决于设备逼真度与环境逼真度。模拟实验中，被试虽知道面临的并不是真实情境，但在设备与环境具有高逼真度的情形下，仍会产生逼真的心理感受，作出如同身临其境的逼真反应。"值得注意的是，心理逼真度在设备和环境逼真度提高到一定程度后，就不再随后二者的提高而增大，因此在人机关系的模拟研究中，如对仿真作过高的要求，就会得不偿失。"①

随着计算机技术的发展，特别是一些虚拟情境制作技术与软件的开发，有许多仿真训练可以采用计算机模拟，虽然逼真度降低了，但付出的成本代价却降低了很多。

---

① 朱祖祥、葛列众、张智君：《工程心理学》，人民教育出版社2000年版，第51—53页。

第十一章

人因工程学研究

阅读材料 11-1

关注人因学

人因学的研究范围很广，它关注的核心是人，要根据人的生理的和心理的能力和极限来设计工作和日常的活动，以避免不安全的、不利于健康的、不舒适的和低效的情况发生。人因学研究的因素很多，比如人的姿势和运动的因素，如坐姿、立姿、提运、推拉等；环境因素，如噪声、震动、照明、小气候环境、化学物品的污染等；信息和操作因素，如信息获得的通道（视觉、听觉、嗅觉等），控制器的特点，显示器与控制器的关系；任务和工作因素，比如，任务应当适当、工作应该有趣。这些因素在很大程度上决定了一个人的工作和日常生活的安全性、健康性、舒适性和有效性。人因学利用来源于多门研究人的科学和技术学科的知识，如人体测量学、生物力学、生理学、心理学、毒物学、机械工程学、工业设计学、信息技术和工业心理学。人因学吸取这些科技领域的知识，并将它们加以整合，运用专门的方法和技术来利用这些知识。人因学与上述这些专门领域的最大区别在于它的跨领域性和应用性。其跨领域性意味着人因学研究人的特征的多个侧面，其应用性意味着人因学要使工作场所、环境及设备适应于人的特点，而不是相反。

人因学有助于解决很多与安全、健康、舒适和效率有关的社会问题。事故是现代生产中很常见的现象，它可以发生在人们的工作中、交通中，甚至家庭日常生活中。有些事故可能导致灾难性后果，如建筑工地的塔吊的倒塌、各种空难、各种矿井的事故都常在新闻报道中看到。在高技术领域，事故的结果则有更加严重的影响，比如美国和前苏联的核电站事故，美国航天飞机的爆炸和俄国潜艇的爆炸事故。在对所有这些事故进行事后的分析时，常常将事故的发生归结于操作人员的操作失误。据分析，在很多情况下，人员的失误只是表面的现象，仔细分析的结果一定会追踪到对于操作者和分配给他们的任务之间的设计的问题。如果在设计过程中充分考虑到人的能力特征和极限水平，就可以大大地减少在工作中和日常生活中发生事故的可能。这就需要人因学研究者们配合国家主管部门制定出有关的人因学标准，以使得设计师们有章可循；在发生法律争执时，也可以有据可依。现在许多发达国家都制定了有关人因学的标准。前不久，

欧洲一些国家提出我国出口到他们国家的产品没有防火功能，不让我们进入他们的市场。这里边固然有这些国家市场保护的考虑，同时也确实有一个提高我们产品的人因学水平的问题。按照人因学的要求，一些可能导致危险的产品，都应设置闭锁装置，以尽可能地保证安全。对于打火机，要做到不懂事的幼儿和宠物不可能拨弄出火的水平。

又比如计算机的屏幕要达到规定的对比度、亮度、字符，要符合一定的规格等等，才能既保证一定的工作效率，又保护使用者的视力。如果我们的企业不注意研究这些标准，我国生产的电脑在进入这些国家市场的过程中，或迟或早将也遇到一些麻烦。

其实，在生活中，特别是人造环境的设计中，产品是否符合人的行为特点往往是性命攸关的。前一些时候我们看到过一些报道，说是在某些失火的情况下，很多受难者都成群地倒在门口，这可能就和门的设计有关。按照人因学的原则，公共场所的出口，一定要设计得在发生人群要冲出去的时候毫无阻碍，也就是说，门要设计得一推就开。这是因为人在慌忙逃生时，一定会本能地向出口拥挤。如果门设计得向内才能打开，人群也就只有堵在门口了。我观察过很多大的公共场所，发现门设计得向里才能打开的还不是少见的。

为了保护工人的安全，我国政府制定了很多人因学标准。比如，为了保证工人的手指不被冲床所伤，在设计上必须要求工人用双手分左右同时按压两个开关才能启动冲床的冲压动作，这样就可以保证不发生致残事故。可是，有的老板为了片面地追求产量，置工人的死活于不顾，竟然私自改动设计，让工人一手送料，一手开动开关，结果造成了很多工人的手指被切断，终身残疾。这是中央电视台去年在焦点访谈节目中报道的。可见，不按人因学的原则和标准办事会导致非常不好的恶果。

交通标志（特别是高速公路的交通标志）的设计同样要考虑符合人因学的要求。有些信息牌上密密麻麻写了很多字，好像给驾车人提供了很多信息。殊不知，在高速行驶中，人在一瞬间能够接受的信息量是有限的，而人又有想对所提供的信息看个究竟的倾向，会不自主地将注意力用于看清楚本来不可能看清楚的信息，而将较少的精力用于驾驶本身，这样一来，本来是好心帮助驾驶者，却反而可能助长事故的发生。与此相类似，为了加快交通，修建了很多立交桥。这本来是好事，但是由于一些桥修得不符合人的习惯（上下出口的位置和方向等），加上给出的信息牌又不能让驾驶人员一目了然，结果是好一点的时候或是使得司

机们动作犹豫,或是走错路;严重的时候,发生事故。

一切产品都是为了人的。如果全社会都来注重人因学在产品设计中的应用,一定会对我们的经济发展有积极的意义。

——资料来源:《产品生产中的人文关怀——关注人因学》(张侃,2002)[①]

阅读材料 11-2

飞机座舱照明与视觉信息显示

飞机座舱照明与飞行信息显示对现代全天候飞机的安全飞行与飞机战斗性能的发挥具有特别重要的作用。在飞机座舱照明与飞行信息显示器的设计中有不少需要工程心理学参与研究的问题。我们曾在这方面做过若干研究。

1. 飞机驾驶舱红、白光照明的心理效应研究

军用飞机驾驶舱的照明必须满足夜间飞行要求。它既要保证飞行员在夜航中能快速而精确地观察各种仪表显示,同时又能保护观察舱外黑暗地空目标的夜视力不受损害。但这两方面的要求存在着矛盾,很难兼顾。因此在对军用飞机的照明研究中,曾存在红光与白光的争论。争论的焦点集中在二者对暗适应或对夜视力的保护作用,以及对信息彩色显示的影响问题上。红光照明有利于保护夜视力,但不利于舱内实现彩色显示。白光照明有利于实现舱内使用彩色编码,但不利于保护夜视力。为了探寻能兼顾舱内外观察要求的飞机照明,我们对红光和不同色温白光的视觉功能进行了较系统的研究,发现人眼的夜视力随照明光色温高低而变化。照明光色温越低,保护夜视力的性能越好,暗适应过程的长短与照明色温高低呈线性关系。发现在亮度较低时,色温 2 000 K 左右的白光照明具有与 620 nm 红光照明相似的夜视力保护作用,用它作飞机驾驶舱仪表照明,兼有红、白光照明的优点。

2. 仪表照明亮度与对比度选择的工效学研究

仪表照明的亮度是影响舱内外观察的重要因素。低亮度有利于保护夜视力,但对辨色、抗强闪光、仪表判读和视觉疲劳等有不利影响。我们研究了仪表

---

① 张侃:《产品生产中的人文关怀——关注人因学》,《科学中国人》,2002(6):20—21。

照明亮度因素与视觉功能的关系，发现1～6 cd/m^{2}亮度范围的照明具有兼顾夜间舱外观察和抗强闪光盲的作用。亮度低于此范围时，会对抗强闪光盲、色标辨别和舒适度感受等产生明显不利影响。若亮度高过此范围，则会对夜视力产生明显不利作用。美、法等航空发达国家将仪表照明均匀度的标准定为3∶1。这个标准要求过高，很难达到。我国航空仪表照明若采用3∶1均匀度标准，将使国产的航空仪表的合格率大为降低。为了制定切合我国实际的仪表照明均匀度标准，我们对仪表照明均匀度与视觉作业工效的关系问题进行了研究。结果证明均匀度5∶1和3∶1具有同等程度的视觉工效。这一研究成果既能满足空军的要求，又能适应我国航空仪表制造的工艺水平，可使我国航空仪表合格率显著提高，生产成本大幅度降低。

3. 不同色温白光差别感觉宽容度的比较研究

对飞机驾驶舱照明的色度要求，一般都以在CIE色度图上规定一个固定的范围来表示。美、英等国规定的航空白区范围有很大差别。我们研究了黑体轨迹上不同色温段白光的色温差别感觉宽容度范围，发现人眼对不同色温段白光的色差感觉阈大小因色温高低不同而发生有规律的不等距变化，表现为人眼对低色温段白光的色度差别感觉阈范围大，随着白光色温提高，色度差别阈限范围缩小。我们根据这一发现，提出CIE色度图的白区范围应按黑体轨迹色温高低作不同限定的见解，即色温从低向高变化时，与之相应的白区范围应由宽向窄变化。根据这一实验结果，我们在CIE色度图黑体轨迹线上作出了6个白光色温差别感觉宽容度范围不同的区域，供飞机白光照明设计者参考。

4. 显示器的亮度、对比度与环境照明匹配的工效学研究

全天候战斗机要求机载显示器的字符亮度和对比度能跟随环境光照强度变化(10^{-1}～10^5 lx)而进行自动调整，这样可以使飞行员不论在黑夜或白天阳光直射时都能看清视觉信息的变化。这就叫显示器的亮度自动控制。实现显示亮度自动控制的关键是要为这种自动控制技术设计者提供人在不同强度背景光照射下正确识别显示字符所要求的亮度和对比度数据。为此，我们开展对电光显示中照明与视觉功能关系的研究。通过实验，我们不仅获得能满足10^{-1}～10^5 lx环境光变化对绿光电光显示的亮度自动控制人机匹配曲线，而且还取得10^{-1}～10^4 lx环境光下的红、橙、黄、蓝、白、黑等多种颜色字符亮度随环境光照强度不同而变化的曲线，为我国航空电光显示器的亮度自动控制系统的研制提供了工程心理学依据。

5. 机载电光显示字符优化的工效学研究

飞行信息大多用符号和文字表示。美、欧不同公司制造的飞机电光显示符号存在较大差别。我国新一代飞机平视显示器和多功能下视显示器将采用汉字符号显示。飞机电光显示器屏幕小，必须控制汉字尺寸和使用汉语简略语。我们对汉字尺寸、汉语简略语和显示符号结构优化等问题进行了研究，获得适用于飞机电光显示器的汉字最佳字体、字大小、字宽度比和笔画宽度等设计参数，选出了多种较优的显示符号。还对飞行员执行飞行任务中常用的汉字语句作多种简略度的比较研究，优选出容易学习、反应速度快、能满足飞行要求的用字最少且航空专用汉语简略语。

6. 灯光信号显示的工效学研究

灯光信号是广泛使用于航空、航海、铁路、公路交通和其他场合的视信息显示形式。室外使用的灯光信号，要求在颜色、亮度和显示方式等方面能受晴、雨、雪、夜等各种环境条件变化的考验，使之在不同环境光影响下都能正确辨认信号的变化。为了满足这一要求，我们曾就环境光照强度变化与颜色灯光信号显示亮度要求的关系问题以及信号显示方式与视觉工效的关系问题进行了研究。获得了相当于夏日和冬日正午阳光直射下辨别红、橙、绿、蓝等不同颜色灯光信号所要求的亮度。发现在强光下正确辨认不同颜色灯光信号需要有不同的信号亮度。还获得与飞机座舱红、白光仪表照明色度、亮度相互兼容的信号适宜亮度。对信号显示方式研究的结果，表明图形显示或图形与文字结合显示方式优于文字显示方式，对红色信号的反应快于白色信号。

7. 汉语语音与视、听综合告警显示的工效学研究

现代飞机告警正在实现向语音告警和多感觉通道综合告警的转变。采用汉语语音告警应是我国新型飞机告警系统的重要特色。我们开展了这方面的研究，获得了有效的语音、灯光、音调三种信号综合的告警方式；汉语语音告警最佳语速参数；传信绩效好的告警短语和故障显示清单用语；多重听觉信号同时出现并发生相互干扰时，为了保证语音告警信号优先被飞行员接收所要求的信号强度参数等研究结果。

——材料来源：《关于照明与视觉显示工效的若干研究》(朱祖祥，2001)①

① 朱祖祥：《关于照明与视觉显示工效的若干研究》，见中国心理学会：《当代中国心理学》，人民教育出版社 2001 年版，第 482—484 页。

建议阅读文献

1. 朱祖祥、葛列众、张智君:《工程心理学》,人民教育出版社 2000 年版,第 1—56 页。
2. 蔡启明、余臻、庄长远:《人因工程》,科学出版社 2005 年版,第 1—15、277—306 页。
3. 阮宝湘、邵祥华:《工业设计人机工程》,机械工业出版社 2005 年版,第 1—30、319—350 页。

复习思考题

1. 如何理解人因工程学、人类工效学、工程心理学、人机系统、人体尺寸、作业疲劳、仿真实验、仿真实验逼真度?
2. 简述人因工程学的发展简史。
3. 简述中国人因工程学发展的现状。
4. 人体参数测量的内容有哪些? 测量的人因学意义是什么?
5. 作业疲劳测量的方法主要有哪些?
6. 仿真实验的逼真度有几个方面?
7. 仿真实验效果的影响因素有哪些?
8. 举例说明人因工程学对于公众日常生活的意义。
9. 你认为人因工程学在中国的发展前景如何?

附录1

常用SPSS程序及其结果选择①

SPSS是Statistics Package for Social Science的英文缩写，为国际上最著名和使用最广泛的统计分析软件，也是教育学、心理学以及其他所有人文学科研究中最有效的资料分析工具。SPSS的统计分析功能十分强大，能完成许多种简单与复杂的资料分析过程。该软件最初是由斯坦福大学三名大学生于1968年开发的，他们基于该系统于1975年在芝加哥合作成立了SPSS公司。该公司不断地对统计分析软件进行改进，先后出现过10余种版本。

20世纪90年代，SPSS公司连续收购了多家同类公司，逐渐由原来单一统计产品的开发与销售转变为向企业、教育科研及政府机构提供全面信息统计决策支持服务，走在了最新流行的“数据仓库”和“数据挖掘”领域前沿的一家综合统计软件公司。公司已经将其英文全称更改为Statistics Product and Service Solutions，意为“统计产品与服务解决方案”。目前，国内大多数心理学本科专业尚未开设SPSS应用课程，但是学生在社会调研、实验教学及毕业论文设计中常常又要依靠该软件进行数据分析。在这一过程中，由于缺乏指导，学生经常会选用不适当的数据分析模块，造成数据分析的错误。为了帮助学生解决这一方面的困难，我们针对心理学实验教学、实验研究中常用的实验研究模式，介绍若干SPSS数据分析程序。学生可根据自己欲分析的数据收集过程与研究模式，选用正确的简明分析程序，快捷地完成数据分析任务。

首先需要说明的是，SPSS软件系统包括三个文件窗口：句法程序编写窗口(syntax)、数据文件窗口(data editor)和结果输出窗口(output)，三个窗口可以方便地通过“Windows”(窗口)菜单条进行切换。我们在下文给出的SPSS的程序均需要写入句法程序编写窗口，然后点击“Run”(运行)菜单条并选择“All”以执行完整程序即可完成相应的数据分析，分析结果放置在结果输出文件窗口中。

① 下文所有的数据分析程序均在SPSS11.01版本中调试完成，所以应尽量在此版本下选用。如果版本不同，程序中可能会出现少数语句不能正常运行的情况。遇到此种情况时，可有如下解决途径：(1)查阅有关软件说明书或国内出版的指导手册；(2)请教有关人士；(3)与本书作者联系和研讨。

在窗口的操作中,除 SPSS 程序自身的命令与菜单外,其他菜单操作与一般的 Windows 应用操作界面相同。

1. 组间设计的平均数差异显著性 t 检验

(1) 实验设计的基本模式

该种实验设计是为了研究某一因素(A)对被试的某种心理状态或行为(R,即因变量)是否产生影响,为此要把这一因素作为自变量(当该因素属于刺激变量或刺激变量引起的机体变量时)或准自变量(当该因素属于被试自身的固有特征时),并选择该因素的两种水平(A_1、A_2)创设两种实验条件。将抽取来的被试随机分成两个独立组,或从两个总体中各随机抽取一个被试组,分别参加 A_1、A_2 条件下的实验,并测量每一被试的某种反应。这就构成了组间实验设计或独立组实验设计。结果分析就是要比较两组被试因变量的平均数是否存在显著性差异。

(2) 组间设计平均数差异显著性的检验程序

```
DATA LIST FREE/A R.
BEGIN DATA.
1 被试数据 2 被试数据
1 被试数据 2 被试数据
……
1 被试数据 2 被试数据
END DATA.
T-TEST GROUPS=A(1, 2)
    /VARIABLES=R.
```


程序说明:

① 程序中将自变量定义为 A,其两个水平 A_1、A_2 分别对应于 A 的变量值 1、2。即当 $A=1$ 时,其代表的实验条件为 A_1;$A=2$ 时,其代表的实验条件为 A_2,因此变量 A 的值将被试实验数据分成了两个组。

② “被试数据”即实验中观测到的每一被试心理或行为反应结果。如果每一被试要进行多次重复测量,则取其平均数代入程序。被试数据的输入不要求确定的顺序,但是必须与其实验条件相应的自变量取值对应。

③ 语句写入时,不能将应有的句号“.”遗漏。

(3) 实例演示及输出结果选择

假如:某心理班 20 名同学参加声、光刺激简单反应时间实验。实验老师将

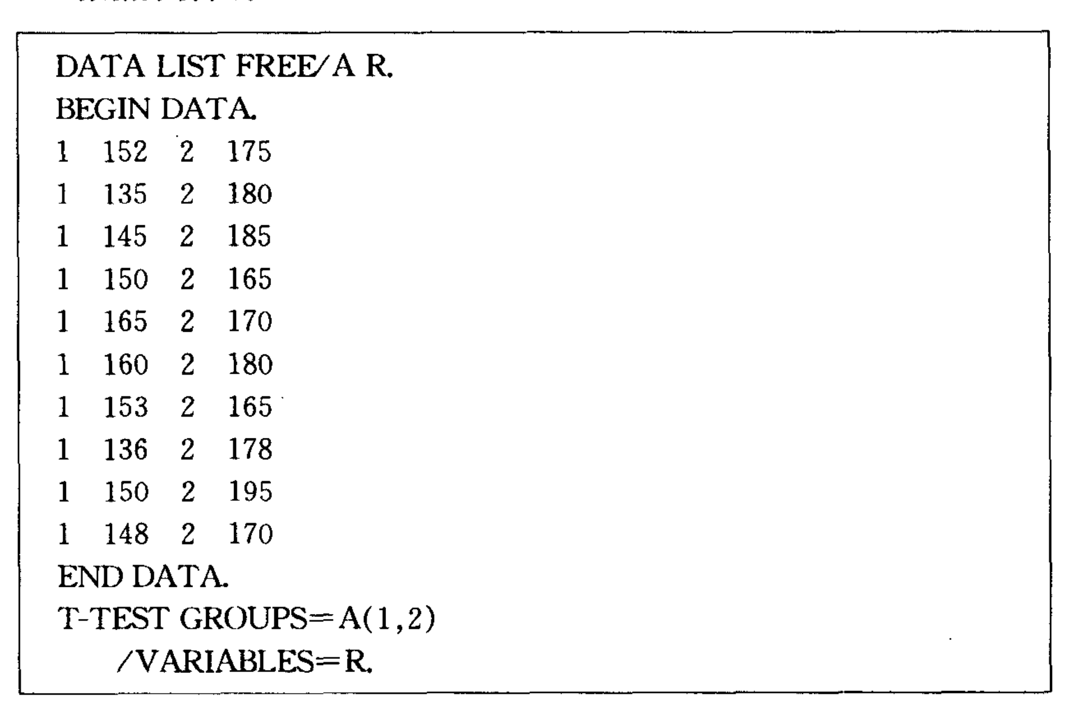
学生随机分成相等的两个组，要求一个组测试声音刺激条件下的反应时间，另一组被试测试灯光刺激条件下的反应时间，结果如表 F1-1 所示。试分析声光刺激条件下，人的简单反应时间是否存在显著性差异？

表 F1-1　声光刺激简单反应时间的比较

被试号	1	2	3	4	5	6	7	8	9	10
声刺激反应时(ms)	152	135	145	150	165	160	153	136	150	148
光刺激反应时(ms)	175	180	185	165	170	180	165	178	195	170

实验中，被试被分成了两个独立组，所以该实验属于组间设计，其平均数差异性检验程序如下所示。程序中 A 代表刺激条件(1 代表“声”、2 代表“光”)、R 代表反应时间。

SPSS 数据分析程序



```
DATA LIST FREE/A R.
BEGIN DATA.
1  152  2  175
1  135  2  180
1  145  2  185
1  150  2  165
1  165  2  170
1  160  2  180
1  153  2  165
1  136  2  178
1  150  2  195
1  148  2  170
END DATA.
T-TEST GROUPS=A(1,2)
    /VARIABLES=R.
```

程序运行结果包括两部分：一部分是描述性结果，另一部分是统计检验结果，如表 F1-2 和表 F1-3 所示。

表 F1-2　Group Statistics

	A	N	Mean	Std. Deviation	Std. Error Mean
R	1.00	10	149.4000	9.33571	2.95221
	2.00	10	176.3000	9.38142	2.96667

表 F1-3 Independent Samples Test(删除了部分多余栏)

	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
方差齐性	.051	.824	−6.427	18	.000
方差不齐性			−6.427	18.000	.000

如表 F1-3,在组间设计的统计检验中输出两个结果,一个是方差齐性时应采用的结果,一个是方差不齐性时应采用的结果。那么,应该选用哪一个结果呢?主要看方差齐性检验的情况,即"Levene's Test for Equality of Variances"栏的结果,如果方差齐性检验统计量 F 比较小,其对应的显著性水平 $p > 0.05$,则方差齐性,就选用该行对应的 t 检验结果;否则,选用"方差不齐性"对应行的 t 检验结果(在方差相等性满足很好时,会出现输出的两个 t 值相等,如本例就是这样的情况)。表 F1-3 显示,该实验中两个组的方差齐性,故选用第一行的结果。将表 F1-2 和表 F1-3 显示的结果汇总,得到该研究数据分析的结果,如表 F1-4 所示。

表 F1-4 声光刺激简单反应时间的比较(组间设计)

声刺激简单反应时间	光刺激简单反应时间	平均数差异量	t	p
149.4±9.3	176.3±9.4	−26.9	−6.427	.000

表 F1-4 所示的结果显示,声光刺激简单反应时间存在非常显著性的差异,说明人们在两种条件下的反应速度有明显不同。

2. 组内设计的平均数差异显著性 t 检验

(1) 实验设计的基本模式

该种实验设计是为了研究某一因素(A)对被试的某种心理状态或行为(R,即因变量)是否产生影响,为此要把这一因素作为自变量,并选择该因素的两种水平(A_1、A_2)创设两种实验条件。将抽取来的被试作为一个被试组参加两种条件下的实验,即每一名被试均参加 A_1、A_2 条件下的实验,并测量每一被试在两种条件下的某种反应。这就构成了组内实验设计或叫做重复测量实验设计。结果分析就是要比较同一组被试在两种条件下测量得到的因变量的平均数是否存在显著性差异。

(2) 组内设计中平均数差异显著性的检验程序

```
DATA LIST FREE/R1 R2.
BEGIN DATA.
被试数据 1   被试数据 2
被试数据 1   被试数据 2
……
被试数据 1   被试数据 2
END DATA.
T-TEST PAIRS=R1 WITH R2.
```

程序说明:

① 程序中将被试在实验条件 A_1 下测得的因变量定义为 R_1,在 A_2 条件下测得的因变量定义为 R_2,然后将每一被试的两个测量值一一对应地输入电脑。

② 这里把每一被试在第一种条件下测得的数据叫"被试数据 1",与该数据对应的被试在另一种条件下测得的数据叫"被试数据 2"。如果每一被试在一种条件下要进行多次测量,则取其平均数代入程序。被试数据的输入不要求确定的顺序,但每一被试的两个数据必须按照相同的前后顺序一一对应地输入,对应关系不能出错。

(3) 实例演示及输出结果选择

假如:某心理班 10 名同学参加声、光刺激简单反应时间实验。实验老师要求每一名同学既要在声音刺激条件下测量简单反应时间,也都要在灯光刺激条件下测量简单反应时间,结果如表 F1-5 所示。试分析声光刺激条件下,人的简单反应时间是否存在显著性差异?

表 F1-5　声光刺激简单反应时间的比较

被试号	1	2	3	4	5	6	7	8	9	10
声刺激反应时(ms)	152	135	145	150	165	160	153	136	150	148
光刺激反应时(ms)	175	180	185	165	170	180	165	178	195	170

我们还使用与表 F1-1 中相同的数据来说明数据分析过程,但是表 F1-5 中数据的采集过程与表 F1-1 不同,因为表 F1-5 中数据来自重复实验设计,即两组数据是由同一组被试测量得到的。该实验属于组内设计,其平均数差异性检验程序如下所示。程序中 R_1 代表声音刺激条件下测得的反应时间、R_2 代表光刺激条件下测得的反应时间(ms)。

SPSS 数据分析程序

```
DATA LIST FREE/R1 R2.
BEGIN DATA.
152  175
135  180
145  185
150  165
165  170
160  180
153  165
136  178
150  195
148  170
END DATA.
T-TEST PAIRS=R1 WITH R2.
```

程序运行结果包括三部分：第一部分是描述性结果，第二部分是相关性分析结果，第三部分是 t 检验结果，如表 F1-6、表 F1-7 和表 F1-8 所示。

表 F1-6 Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	R_1	149.400 0	10	9.335 71	2.952 21
	R_2	176.300 0	10	9.381 42	2.966 67

表 F1-7 Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	R1 & R2	10	−.259	.470

表 F1-8 Paired Samples Test

		Paired Differences	t	df	Sig. (2-tailed)
Pair 1	R_1-R_2	−26.900 0	−5.728	9	.000

根据上述输出的结果，可以将数据分析的结果总结为表 F1-9 所示的形式。

表 F1-9 声光刺激简单反应时间的比较(组内设计)

声刺激简单反应时间	光刺激简单反应时间	平均数差异量	t	p
149.4±9.3	176.3±9.4	−26.9	−5.728	.000

表 F1-9 所示的结果显示，声光刺激简单反应时间存在非常显著性的差异，说明人们在两种条件下的反应速度有明显不同。

3. 单因素方差分析(独立组设计)

(1) 实验设计的基本模式

该种实验设计是为了研究某一因素(A)对被试的某种心理状态或行为(R,即因变量)是否产生影响，为此要把这一因素作为自变量(当该因素属于刺激变量或刺激变量引起的机体变量时)或准自变量(当该因素属于被试自身的固有特征时)，并选择该因素的三种或三种以上的水平(A_1、A_2、A_3……)创设超过两种以上的实验条件。将抽取来的被试随机分成三个以上的独立组，或从三个以上的总体中各随机抽取一个被试组，分别参加 A_1、A_2、A_3……条件下的实验，并测量每一被试的某种反应 R。这就构成了单因素完全随机实验设计。超过了两个独立组的实验数据的分析不能直接使用独立组的两两 t 检验，而是要使用单因素(独立组)方差分析程序，即“ONEWAY ANOVA”①分析程序。使用这一程序，不仅可得到各独立组测试结果的平均数、标准差等描述性结果，而且可以进行组间方差齐性检验、自变量影响效应检验、事后的多重比较(独立组的两两比较)。

(2) 单因素(独立组)方差分析程序

```
DATA LIST FREE/A R.
BEGIN DATA.
1  被试数据  2  被试数据  3  被试数据
1  被试数据  2  被试数据  3  被试数据
……
1  被试数据  2  被试数据  3  被试数据
END DATA.
ONEWAY R BY A
  /STATISTICS DESCRIPTIVES HOMOGENEITY
  /POSTHOC=LSD ALPHA(.05).
```

程序说明：

① 程序中将自变量定义为 A，其有三个水平 A_1、A_2、A_3 分别对应于 A 的变量值 1、2、3(如果超过三个水平，则按照相似的方法增加数据列)。当 $A=1$ 时，其代表的实验条件为 A_1；$A=2$ 时，其代表的实验条件为 A_2；$A=3$ 时，其代表的实验条件为 A_3，因此变量 A 的值将被试实验数据分成了三个组。

① ANOVA 是“方差分析”(Analysis of Variance)的英文缩写。

② “被试数据”即实验中观测到的每一被试心理或行为反应结果。如果每一被试要进行多次测量，则取其平均数代入程序。被试数据的输入不要求确定的顺序，但是必须与其实验条件相应的自变量取值对应。

(3) 方差分析命令语句“ONEWAY R BY A/STATISTICS DESCRIPTIVES HOMOGENEITY /POSTHOC=LSD ALPHA(.05).”包括三个分语句：“ONEWAY R BY A”是核心语句，其执行的结果是输出方差分析表；“/STATISTICS DESCRIPTIVES HOMOGENEITY”是要求程序执行方差分析时同时输出各被试组测量数据的主要统计量，以及对各组数据的方差奇性进行检验，输出结果主要包括各组数据的平均数、标准差等，以及方差齐性检验结果；“/POSTHOC=LSD ALPHA(.05).”则是要求程序在进行方差分析之后，进行独立组之间的两两比较，检验两两之间平均数的差异显著性（达到显著性的标准是 $p < 0.05$），也叫做多重比较或事后多重比较(post hoc)。

(3) 实例演示及输出结果选择

假如：某研究者想考察初一、初三、高二学生的元认知监控能力水平是否存在差异，于是从某校三个年级学生中各随机抽取 10 名男生，并对他们进行了元认知监控能力的测试，结果如表 F1-10 所示。试问，三个年级间学生的监控能力是否具有显著差异？

表 F1-10　中学生元认知监控能力的发展性研究

被试号	1	2	3	4	5	6	7	8	9	10
初一学生数据	42	25	35	40	55	50	43	26	40	38
初三学生数据	65	70	75	55	60	70	55	68	65	60
高二学生数据	75	80	85	65	70	80	65	78	95	70

实验中，被试来自三个不同年级，年级为准自变量 GRADE，该实验属于单因素完全随机实验设计。因为被试组超过了两个，其平均数差异性检验需要采用方差分析程序，即“ONEWAY ANOVA”分析程序，如下所示。程序中 GRADE 代表准自变量“学生年级”，(1 代表“初一”、2 代表“初三”、3 代表“高二”)、SCORE 代表反应“元认知监控能力测试分数”。

SPSS 数据分析程序

```
DATA LIST FREE/GRADE SCORE.
BEGIN DATA.
```

(续表)

```
1  42  2  65  3  75
1  25  2  70  3  80
1  35  2  75  3  85
1  40  2  55  3  65
1  55  2  60  3  70
1  50  2  70  3  80
1  43  2  55  3  65
1  26  2  68  3  78
1  40  2  65  3  95
1  38  2  60  3  70
END DATA.
ONEWAY SCORE BY GRADE
  /STATISTICS DESCRIPTIVES HOMOGENEITY
  /POSTHOC=LSD ALPHA(.05).
```

程序运行结果包括四部分：第一部分是各组数据的描述性分析结果（平均数、标准差、标准误等）；第二部分是各组数据间方差相等性的检验结果；第三部分是自变量或准自变量效应的方差分析结果；第四部分是方差分析之后进行的多重比较，即独立组间的两两比较。这些结果分别见表 F1-11、表 F1-12、表 F1-13和表 F1-14 所示。

表 F1-11 Descriptives

	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
1.00	10	39.400 0	9.335 71	2.952 21	25.00	55.00
2.00	10	64.300 0	6.700 75	2.118 96	55.00	75.00
3.00	10	76.300 0	9.381 42	2.966 67	65.00	95.00
Total	30	60.000 0	17.681 08	3.228 11	25.00	95.00

从表 F1-11 显示的信息，可以描述三个被试组因变量的平均水平、标准差等基本信息。就本例来说，这些信息反映了学生元认知监控能力的发展趋势（需要提醒的是，这里的数据是假想的，所以数据显示的结果也非研究结论）。

表 F1-12 Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
.349	2	27	.708

表 F1-12 显示的信息主要是关于各被试组测试结果的方差相等性，这是方

差分析之前必须采取的数据分析过程，因为方差分析的必要前提是方差齐性。如表 F1-12 显示，各组间的方差没有明显差异（Levene Statistic =. 349, $p<.708$），即方差齐性，符合方差分析的要求。

表 F1-13　ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7 085.400	2	3 542.700	48.295	.000
Within Groups	1 980.600	27	73.356		
Total	9 066.000	29			

表 F1-13 显示的是方差分析结果，信息包括变异源、变异平方和（sum of squares）、自由度（df）、均方或称方差（mean square，简称 MS）、检验统计量 F、平均数差异显著性（sig）。本例中，自变量效应（Between Groups）检验中 $F=48.295$，显著性水平 $p=.000<.001$。

表 F1-14　Multiple Comparisons LSD

(I) GRADE	(J) GRADE	Mean Difference (I-J)	Std. Error	Sig.
1.00	2.00	−24.900 0	3.830 29	.000
	3.00	−36.900 0	3.830 29	.000
2.00	1.00	24.900 0	3.830 29	.000
	3.00	−12.000 0	3.830 29	.004
3.00	1.00	36.900 0	3.830 29	.000
	2.00	12.000 0	3.830 29	.004

* The mean difference is significant at the .05 level.

表 F1-14 显示的信息则是独立组间平均数的两两比较，以考察平均数差异的具体表现。如本例中，三个年级间的元认知监控能力均存在两两间的显著性差异。

4. **单因素方差分析（重复测量设计）**

(1) 实验设计的基本模式

该种实验设计是为了研究某一因素（A）对被试的某种心理状态或行为（SCORE，即因变量）是否产生影响，为此把这一因素作为自变量，并选择该因素的三种或三种以上的水平（A_1、A_2、A_3……）创设超过两种以上的实验条件。将抽取来的被试作为一个被试组，每个被试均参加 A_1、A_2、A_3……条件下的实

验，并测量每一被试在每一种实验条件下的某种反应，得到每一被试的观测分数 SCORE1、SCORE2、SCORE3……这就构成了单因素重复测量的实验设计。超过两组实验数据的分析不能直接使用重复测量的两两 t 检验，而要使用单因素(重复测量)方差分析程序，即以“GLM”来引导分析程序。①使用这一程序，可以得到各条件下测试结果的平均数、标准差等描述性统计结果，以及自变量影响效应的检验结果。

(2) 单因素(重复测量)方差分析程序

```
DATA LIST FREE/SCORE1 SCORE2 SCORE3.
BEGIN DATA.
被试数据 1   被试数据 2   被试数据 3
被试数据 1   被试数据 2   被试数据 3
……
被试数据 1   被试数据 2   被试数据 3
END DATA.
GLM SCORE1 SCORE2 SCORE3
  /WSFACTORS=SCORE(3)
  /PRINT=DESCRIPTIVES.
```

程序说明：

① 程序中的变量 SCORE1、SCORE2 和 SCORE3 分别对应于一组被试在三种不同实验条件下的观测分数(如果自变量超过三个水平，则按照相似的方法增加观测数列)。

② 在命令语句“GLM SCORE1 SCORE2 SCORE3/WSFACTORS=SCORE(3)/PRINT = DESCRIPTIVES.”中：“GLM SCORE1 SCORE2 SCORE3/WSFACTORS=SCORE(3)”是该程序的核心，其执行的结果是输出方差分析表，它意指对三列观测数据进行重复测量的方差分析，并用 WSFACTORS=SCORE(3)说明了重复测量变量的名称和水平数；“/PRINT=DESCRIPTIVES”则是为了程序执行中能够输出每一数据单元的基本统计量平均数和标准差等。

(3) 实例演示及输出结果选择

假如：为研究刺激光的颜色对反应时间有无明显的影响，研究者抽取 10 名大学生进行不同色光刺激条件下的选择反应时测定。实验采用单因素重复实验

① GLM 是“General Linear Model”的英文缩写。在多因素方差分析中，如果假定各变异源引起的数据变异具有简单相加关系，则将其称为“一般线性模型”，即 GLM。

设计，即每一被试均完成四种颜色（红、黄、绿、蓝）光刺激下的选择反应时测定。得到如表 F1-15 所示的数据。

表 F1-15　光刺激的颜色对选择反应时的影响

被试编号	RT1(红)	RT2(黄)	RT3(绿)	RT4(蓝)
1	350	412	400	412
2	300	350	320	380
3	380	380	360	350
4	365	450	421	430
5	340	350	370	400
6	386	430	400	413
7	300	285	300	360
8	400	465	412	430
9	312	360	350	380
10	360	425	395	410

此实验设计中不同色光下反应时间的差异比较需要借助方差分析来进行。由于实验采用组内设计，所以方差分析要用 GLM 命令。具体的程序如下：

```
DATA LIST FREE/RT1 TO RT4.
BEGIN DATA.
350  412  400  412
300  350  320  380
380  380  360  350
365  450  421  430
340  350  370  400
386  430  400  413
300  285  300  360
400  465  412  430
312  360  350  380
360  425  395  410
END DATA.
GLM RT1 RT2 RT3 RT4
  /WSFACTORS=RT(4)
  /PRINT=DESCRIPTIVES.
```


程序运行输出结果主要包括两部分：第一部分是关于各条件下观测数据的平均数、标准差等描述性结果；第二部分主要是检验被试内因子的方差分析结果，即以“Tests involving ‘RT’ Within-Subject Effect.”为标题的方差分析结果，如表 F1-16 和表 F1-17 所示。

表 F1-16 Cell Means and Standard Deviations

Variable	Mean	Std. Dev.	N
RT1	349.300	35.839	10
RT2	390.700	55.532	10
RT3	372.800	40.293	10
RT4	396.500	27.870	10

表 F1-17 Tests involving 'RT' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
RT	13 524.48	3	4 508.16	11.00	.000
WITHIN CELLS	110 69.77	27	409.99		

从本例数据结果看,不同色光刺激下被试的反应时间存在显著性差异($F=11.00$, $p=.000$)。

5. *多因素方差分析(独立组设计)*

(1) 实验设计的基本模式

该种实验设计是为了研究两个以上的因素(A、B、C ……)对被试的某种心理状态或行为(R,即因变量)是否产生影响,以及这些影响因素之间是否存在交互作用,为此要在研究过程中操纵或改变两个以上的自变量(当该因素属于刺激变量或刺激变量引起的机体变量时)或准自变量(当该因素属于被试自身的固有特征时),而且每一研究因子都有两种或两种以上的水平(A_1、A_2、A_3……;B_1、B_2、B_3……),这样就创设出一系列不同的实验条件或划分出具有多种不同特征的被试组(是指用多个准自变量区分出的被试组,如:大一男生、大三男生、大一女生、大三女生,就是由两个准自变量区分出的四个被试组),然后每一组被试只在一种实验处理下接受因变量的测量。在这种实验设计中,涉及两个以上的自变量或准自变量,数据分析不能直接使用 t 检验,也不能直接使用单因素方差分析,而要使用多因素完全随机实验的方差分析程序,即"UNIANOVA"分析程序。使用这一方差程序,不仅可以得到各独立组被试测试结果的平均数、标准差等描述性统计结果,而且可以进行组间的方差齐性检验、自变量影响主效应的检验、变量间交互效应检验、事后的多重比较(独立组的两两比较)等。

(2) 多因素完全随机实验的方差分析程序

SPSS 程序(以 3×3 实验设计为例)

```
DATA LIST FREE/A B SCORE.
BEGIN DATA.
1  1  被试数据
……
1  1  被试数据
1  2  被试数据
……
1  2  被试数据
1  3  被试数据
……
1  3  被试数据
2  1  被试数据
……
2  1  被试数据
2  2  被试数据
……
2  2  被试数据
2  3  被试数据
……
2  3  被试数据
3  1  被试数据
……
3  1  被试数据
3  2  被试数据
……
3  2  被试数据
3  3  被试数据
……
3  3  被试数据
END DATA.
UNIANOVA SCORE BY A B
  /POSTHOC=A B (LSD)
  /PRINT=DESCRIPTIVE HOMOGENEITY.
```


程序说明:

① 这一程序是针对 3×3 完全随机实验设计的,如果有更多的变量或更多变量水平,则按照类似方法增加数据行和数据列。

② “被试数据”即实验中观测到的每一被试心理或行为反应结果。如果每一被试要进行多次重复测量,则取其平均数代入程序。被试数据的输入不要求确定的顺序,但必须与其实验条件相应的自变量取值对应。

③ 在命令语句“UNIANOVA SCORE BY A B /PRINT = DESCRIPTIVE HOMOGENEITY /POSTHOC= A B (LSD)”中包括三个分语句：“UNIANOVA SCORE BY A B”是该语句的核心，其执行的结果是输出方差分析表，这里的“UNIANOVA”是指“单因变量的方差分析程序”之意，它执行对某一因变量的方差分析；“/PRINT = DESCRIPTIVE HOMOGENEITY.”是要求程序在进行方差分析的同时输出各被试组测量数据的主要统计量，以及对各组数据的方差相等性进行检验，输出结果主要包括：各组数据的平均数、标准差，以及方差齐性检验结果；“/POSTHOC= A B (LSD)”是方差分析之后的多重比较，即对实验中各独立组之间的差异性做两两比较。

(3) 实例演示及输出结果选择

假设：某研究者想探明一“提问单训练”程序对学生解题监控能力发展的影响，同时考虑这种影响可能与学生已有的学习成绩有关。于是，研究者设计了一个 3×3 的实验，其中的两个自变量或准自变量分别是训练方法和学生现有学习成绩。训练方法包括三个水平：自我提问训练、他人提问训练、不训练(控制组)；学生现有学习成绩分优、中、差三个水平。经过一定的训练周期后(控制组不参加训练)，参加解题监控能力与监控意识测试，结果如表 F1-18 所示。这一研究数据的方差分析要考察训练方式(自变量 A)和学生现有成绩水平(准自变量 B)对解题监控能力发展的主效应及交互效应。

表 F1-18 学习成绩不同的学生接受“提问单训练”的效果比较(每一单元被试 5 人)

学习成绩 \ 训练方式	不训练					他人提问训练					自我提问训练				
优 等	85	70	80	88	85	85	80	89	90	88	90	80	85	95	90
中 等	65	70	60	65	60	70	75	65	60	70	85	80	70	80	85
差 等	50	45	55	40	50	50	65	60	55	60	75	80	85	80	70

SPSS 数据分析程序

```
DATA LIST FREE/A B SCORE.
BEGIN DATA.
1  1  85
1  1  70
1  1  80
1  1  88
```

(续表)

```
1  1  85
1  2  65
1  2  70
1  2  60
1  2  65
1  2  60
1  3  50
1  3  45
1  3  55
1  3  40
1  3  50
2  1  85
2  1  80
2  1  89
2  1  90
2  1  88
2  2  70
2  2  75
2  2  65
2  2  60
2  2  70
2  3  50
2  3  65
2  3  60
2  3  55
2  3  60
3  1  90
3  1  80
3  1  85
3  1  95
3  1  90
3  2  85
3  2  80
3  2  70
3  2  80
3  2  85
3  3  75
3  3  80
3  3  85
3  3  80
3  3  70
END DATA.
UNIANOVA SCORE BY A B
  /POSTHOC =A B ( LSD )
  /PRINT = DESCRIPTIVE HOMOGENEITY.
```

程序运行结果包括四部分：各实验单元观测数据的描述性结果、方差齐性检验结果、方差分析表、针对两个自变量或准自变量各自进行的多重比较，分别如表F1-19、表F1-20、表F1-21和表F1-22所示。

表F1-19 Descriptive Statistics

A	B	Mean	Std. Deviation	N
1.00	1.00	81.600 0	7.092 25	5
	2.00	64.000 0	4.183 30	5
	3.00	48.000 0	5.700 88	5
2.00	1.00	86.400 0	4.037 33	5
	2.00	68.000 0	5.700 88	5
	3.00	58.000 0	5.700 88	5
3.00	1.00	88.000 0	5.700 88	5
	2.00	80.000 0	6.123 72	5
	3.00	78.000 0	5.700 88	5

表F1-19所示的结果显示出各实验单元被试测试分数的平均数和标准差。

表F1-20 Levene's Test of Equality of Error Variances

F	df1	df2	Sig.
.241	8	36	.980

表F1-20所示的结果，方差齐性的条件成立（$F=.241$，$p=.980$），可以进行方差分析。

表F1-21 Tests of Between-Subjects Effects

Source	Type Ⅲ Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7 528.711	8	941.089	29.813	.000
Intercept	236 168.889	1	236 168.889	7 481.591	.000
A	2 348.978	2	1 174.489	37.207	.000
B	4 391.111	2	2 195.556	69.553	.000
A * B	788.622	4	197.156	6.246	.001

(续表)

Source	Type Ⅲ Sum of Squares	df	Mean Square	F	Sig.
Error	1 136.400	36	31.567		
Total	244 834.000	45			
Corrected Total	8 665.111	44			

表 F1-21 所示的结果是方差分析的主要结果，显示出自变量的主效应和交互效应，如自变量 A 的主效应达到了非常显著性的水平 (FA = 37.207, p =.000)。

表 F1-22 Multiple Comparisons LSD

	(I) A	(J) A	Mean Difference (I-J)	Std. Error	Sig
A	1.00	2.00	−6.266 7	2.051 56	.004
		3.00	−17.466 7	2.051 56	.000
	2.00	1.00	6.266 7	2.051 56	.004
		3.00	−11.200 0	2.051 56	.000
	3.00	1.00	17.466 7	2.051 56	.000
		2.00	11.200 0	2.051 56	.000
	(I)B	(J)B	Mean Difference (I-J)	Std. Error	Sig
B	1.00	2.00	14.666 7	2.051 56	.000
		3.00	24.000 0	2.051 56	.000
	2.00	1.00	−14.666 7	2.051 56	.000
		3.00	9.333 3	2.051 56	.000
	3.00	1.00	−24.000 0	2.051 56	.000
		2.00	−9.333 3	2.051 56	.000

表 F1-22 所示的结果是自变量各水平间因变量平均数的差异性比较，即两两比较。

6. 多因素方差分析(重复测量设计)

(1) 实验设计的基本模式

该种实验设计也是为了研究两个以上的因素(A、B、C……)对被试的某种心理状态或行为(R,即因变量)是否产生影响，以及这些影响因素之间是否存在

交互作用，为此要在研究过程中操纵或改变两个以上的自变量，而且每一自变量都有两种或两种以上的水平（A_1、A_2、A_3……；B_1、B_2、B_3……），这样就创设出一系列不同的实验条件。然后所有被试均接受所有实验条件下的测量得到一系列因变量值。在这种实验设计中，涉及两个以上的自变量，数据分析不能直接使用 t 检验，也不能直接使用单因素方差分析，而要使用多因素重复实验的方差分析程序。使用这一程序，不仅可以得到各实验条件下被试测试结果的平均数、标准差等描述性结果，而且可以检验自变量影响的主效应、交互效应等，进一步的事后多重比较则可以对不同实验条件作两两比较。

（2）多因素重复测量的方差分析程序

SPSS 程序（以 2×3 实验设计为例）

```
DATA LIST FREE/A1B1 A1B2 A1B3 A2B1 A2B2 A2B3.
BEGIN DATA.
被试数据  被试数据  被试数据  被试数据  被试数据  被试数据
被试数据  被试数据  被试数据  被试数据  被试数据  被试数据
……
被试数据  被试数据  被试数据  被试数据  被试数据  被试数据
END DATA.
GLM A1B1 A1B2 A1B3 A2B1 A2B2 A2B3
  /WSFACTOR = A2B3
  /PRINT = DESCRIPTIVE.
```

程序说明：

① 这一程序是针对 2×3 重复测量实验设计的，如果有更多变量或更多变量水平，则按照类似的方法增加数据列。

② 变量名是由自变量各水平编码符号组合而成，它代表了某一组合的实验条件。如 A_1B_2 代表了自变量 A 的水平 1 和自变量 B 的水平 2 结合而成的一种实验条件，该列被试数据就是所有被试在这一条件下测量得到的因变量值。

③ 这一程序的命令语句"GLM A1B1 A1B2 A1B3 A2B1 A2B2 A2B3/WSFACTOR=A2B3/PRINT= DESCRIPTIVE."中包括三个分语句："GLM A1B1 A1B2 A1B3 A2B1 A2B2 A2B3"是该语句的核心，其定义了该方差分析程序所要分析的测量数列，"WSFACTOR=A2 B3"是说明程序中有两个重复测量的自变量，且分别有 2 个和 3 个水平；"PRINT=DESCRIPTIVE."则是要求程序在执行方差分析的同时，输出各测量单元的描述性分析结果，包括平均数和标准差等。

（3）实例演示及输出结果选择

假设:某研究者想探明一"提问单训练"程序对学生语文、数学、物理学科学习监控绩效的影响。于是,研究者设计了一个2×3的实验,其中的两个自变量分别是训练和学科门类。训练分训练前和训练后两个水平;学科门类涉及语文、数学、物理三个学科。研究者选取了5名被试,先对他们进行语文、数学、物理学习监控水平测试,然后开始一定时段的训练,训练结束后再对学生进行语文、数学和物理学习监控水平测试,结果如表F1-23所示。

表 F1-23 训练前后学生不同课程学习监控绩效测量

学科 \ 训练方式	训练前测试成绩					训练后测试成绩				
语 文	85	70	80	88	85	85	80	89	90	88
数 学	65	70	60	65	60	70	75	65	60	70
物 理	50	45	55	40	50	50	65	60	55	60

SPSS 数据分析程序

```
DATA LIST FREE/A1B1 A1B2 A1B3 A2B1 A2B2 A2B3.
BEGIN DATA.
85  65  50  85  70  50
70  70  45  80  75  65
80  60  55  89  65  60
88  65  40  90  60  55
85  60  50  88  70  60
END DATA.
GLM A1B1 A1B2 A1B3 A2B1 A2B2 A2B3
  /WSFACTOR = A2 B3
  /PRINT = DESCRIPTIVE.
```


程序运行结果主要包括两部分:各实验单元观测数据的描述性分析结果、方差分析表,分别如表F1-24、表F1-25所示。

F1-24 Descriptive Statistics

	Mean	Std. Deviation	N
A_1B_1	81.600 0	7.092 25	5
A_1B_2	64.000 0	4.183 30	5
A_1B_3	48.000 0	5.700 88	5
A_2B_1	86.400 0	4.037 33	5

（续表）

	Mean	Std. Deviation	N
A_2B_2	68.000 0	5.700 88	5
A_2B_3	58.000 0	5.700 88	5

表 F1-24 显示的结果是各实验单元的测量分数的平均数与标准差。

F1-25 Tests of Within-Subjects Effects(删除了一些多余信息)

Source		Sum of Squares	df	Mean Square	F	Sig.
A	Sphericity Assumed	294.533	1	294.533	13.678	.021
Error(A)	Sphericity Assumed	86.133	4	21.533		
B	Sphericity Assumed	4 846.667	2	2 423.333	40.277	.000
Error(B)	Sphericity Assumed	481.333	8	60.167		
A * B	Sphericity Assumed	53.067	2	26.533	1.535	.273
Error(A * B)	Sphericity Assumed	138.267	8	17.283		

表 F1-25 所示的结果给出了自变量 A、B 的主效应的显著性水平，以及二者的交互效应的显著性水平。如本例中，A 和 B 的主效应显著，但二者的交互效应不显著。

在此需要特别说明两点：

① 如果方差分析中包含协变量，则需要将协变量的变量名加在变量名定义行里(即程序的第一行)，并在方差分析的主命令中加上以“WITH”连接的协变量名。如在多因素重复实验设计方差分析的主命令“GLM A_1B_1 A_1B_2 A_1B_3 A_2B_1 A_2B_2 A_2B_3”中加上“WITH IQ”，使之变为“GLM A_1B_1 A_1B_2 A_1B_3 A_2B_1 A_2B_2 A_2B_3 WITH IQ”，该程序在方差分析中就会将“IQ”作为协变量引入分析过程。

② 在方差分析中，只要包含有重复测量的自变量，就需要使用 GLM 开头的主命令，它引导程序完成有重复测量自变量、协变量的数据的方差分析过程。

7. **独立性卡方检验**①

独立性卡方检验是一种用于评估两个测量称名量表或顺序量表的变量之间关系的假设检验方法，或者用于评估两个独立组中比例关系差异性的假设检验。

① ［美］弗雷德里克・J・格拉维特、罗妮安・B・福泽诺著，邓铸、姜子云、蒋小慧等译：《行为科学研究方法》，陕西师范大学出版社 2005 年版，第 303—304 页。

我们使用下列数据来演示独立性卡方检验的过程，这些数据反映了一个 200 人样本中的频数分布。根据性格（内向和外向）和颜色偏爱（红、黄、绿、蓝）两个变量对每个人进行归类，每个单元格中的数字代表了相应性格特征和颜色偏爱的人的数量，如内向且喜爱红色的有 10 人。这里的频数值被称为是观测频数或 F_O 值。

颜色偏好

	红	黄	绿	蓝	
内向	10	3	15	22	50
外向	90	17	25	18	150
	100	20	40	40	

针对这样的数据，零假设的设定有两种方式：性格与颜色偏好没有关系；对于内向与外向性格的人来说，颜色偏好的分布是相同的。对其检验的步骤如下。

第一步

卡方检验的第一步是计算一个假设的频数集合，这个频数集合反映了当样本中的频数分布与零假设完全一致时的情形。这些假设的频数被称为是期望频数（expected frequencies）或 f_E 值。对于矩阵中的每一单元格来说，其期望频数可以通过下列方法计算：

$$f_E = (\text{所在行总数} \times \text{所在列总数}) / \text{总数量}$$

比如矩阵中最上边一行左边第一个单元格，所在行是第一行，其总数为 50；所在列是第一列，其总数为 100；在整个研究中被试的总数是 200 人，所以这个单元格中的期望频数为：

$$f_E = (50 \times 100)/200 = 25$$

下边的矩阵就是一个完全的期望频数集合。

颜色偏好

	红	黄	绿	蓝	
内向	25	5	10	10	50
外向	75	15	30	30	150
	100	20	40	40	

第二步

卡方检验的第二步是计算卡方值（χ^2），它提供了一种对观测到的频数分布

(实际数据)与期望分布(假设)差异性的测量。卡方值的计算公式为:

$$\chi^2 = \sum (f_O - f_E)^2 / f_E$$

根据我们例子中的数据,一步一步的计算如下:

a. 求出矩阵中每一单元格的期望频数与观测频数之间的差异;

b. 将差异值平方;

c. 用期望频数去除差异量的平方值;

d. 求每一类别中上述计算得到的结果的总和。

f_O	f_E	$(f_O - f_E)$	$(f_O - f_E)^2$	$(f_O - f_E)^2/f_E$
10	25	15	225	9.00
3	5	2	4	0.80
15	10	5	25	2.50
22	10	12	144	14.40
90	75	15	225	3.00
17	15	2	4	0.27
25	30	5	25	0.83
18	30	12	144	4.80

$$\chi^2 = \sum (f_O - f_E)^2 / f_E = 35.60$$

查 χ^2 分布检验表来确定 $\chi^2 = 35.60$ 是不是大到足以证明统计学上的显著性。卡方统计量的自由度按以下公式计算: $df = (C_1 - 1)(C_2 - 1)$。

其中 C_1 是第一个自变量的分类数, C_2 是第二个自变量的分类数。就我们的数据来说, $df = (2-1)(4-1) = 3$。

当 $df = 3$ 时, $\alpha < 0.01$ 显著性水平所要求的最小卡方值是 $\chi^2 = 11.34$。本例中得到的卡方值超过这一标准,所以根据我们使用的零假设,可以得到结论:

a. 性格与颜色偏好存在显著性关系;或

b. 内向性格人的颜色偏好与外向性格的人有显著不同。

那么,这一例子中数据分析的 SPSS 程序是如何构成的呢?

独立性卡方检验的 SPSS 程序:

```
DATA LIST FREE/A B NUMBER.
BEGIN DATA.
1 1 10
1 2 90
```

（续表）

```
2 1 3
2 2 17
3 1 15
3 2 25
4 1 22
4 2 18
END DATA.
WEIGHT BY NUMBER.
CROSSTABS
  /TABLES = B BY A
  /FORMAT = AVALUE TABLES
  /STATISTIC = CHISQ
  /CELLS = COUNT.
```

程序说明：

① 该程序中的变量 *A* 和 *B* 是被试的分类变量，而 NUMBER 则是各个类别中被试的人数。

② 程序中包括两个命令语句。第一个命令语句为"WEIGHT BY NUMBER."是将 NUMBER 的变量值加权，这样，NUMBER 的变量值就代表着人数，而不是代表一个个案的观测值。

③ 程序中的第二个命令语句"CROSSTABS /TABLES＝B BY A /FORMAT＝AVALUE TABLES /STATISTIC＝CHISQ /CELLS＝COUNT"主要是对两个自变量在被试分类上是否相互独立进行检验。该命令通过计算卡方值来检验两个变量独立性假设，而且是卡方值越大，两个变量的相互依存性越明显，即两个变量的交互作用越明显。

程序运行输出的结果主要包括两部分：(1)第一部分是数据交叉列联表，显示被试人数在两个变量不同水平上的分布，如表 F1-26 所示；(2)第二部分是两个

表 F1-26　B * A Crosstabulation

		A				Total
		1	2	3	4	
B	1	10	3	15	22	50
	2	90	17	25	18	150
Total		100	20	40	40	200

变量独立性的卡方检验，结果如表 F1-27 所示，结果显示，本例中的两个自变量具有显著的相互依存性（$\chi^2 = 35.60$, $p = .000$），与前文利用公式计算的结果完全一致。

表 F1-27　Chi-Square Tests

	Value	df	Sig.
Pearson Chi-Square	35.600	3	.000
Likelihood Ratio	35.033	3	.000
Linear-by-Linear Association	34.476	1	.000
N of Valid Cases	200		

如需用其他方差分析程序，请查阅相关统计学和 SPSS 应用教程。

附录 2

E-Prime 在心理学实验中的应用

20 世纪 90 年代以来，随着计算机软件技术的迅速发展，心理学的实验教学和实验研究出现了计算机化趋势，即通过实验软件或计算机程序，实现心理学实验过程的计算机控制和实验结果的记录与分析。早在 1971 年，美国心理学界就成立了计算机的心理学应用学会(Society for Computers in Psychology)，以满足心理学研究和教学中计算机应用的讨论和信息交换的需要。当前，计算机程序控制的实验已被广泛应用于心理学研究和教学的各分支领域，如认知神经科学、感知、记忆、思维等。Psychonomic 学会出版的期刊 *Behavior Research Methods, Instruments and Computers* 给予心理学研究中的计算机应用问题以特别的关注。近几年，国内心理学工作者也已经意识到计算机在心理学实验教学和研究中的巨大作用，研发了多种心理学实验软件，但功能还主要限于心理学的实验教学演示，研究人员则要花费大量时间学习汇编语言，才能编制出所需要的实验控制程序。

迄今为止，国内尚未开发出真正意义上的、成熟的心理学实验生成系统，所用实验程序大多会请专业编程人员使用 VB 语言、C 语言等编制。就国内心理学实验研究和教学而言，研究者编制实验程序时主要存在以下问题：(1)研究者请专业编程人员编制实验程序，一方面，要花费大量时间向编程人员解释实验意图，另一方面，需要投入大量的编程费用，可操作性不强；(2)具有一定编程基础的研究者自己编制实验常常面临技术上的难题；(3)更多的编程者使用 Visual Basic 来编写实验控制程序，其时间精度不够、误差较大、数据分析和处理存在许多缺陷，其他编程语言则更难掌握。

总之，标准化的心理实验生成系统在心理学教学和研究中显得十分重要。这种系统必须简单易用，不会耗费研究人员太多的时间，能够较好地解决计时精度、系统误差等问题，能降低数据处理的难度，便于在研究中推广。作为一种标准化的心理学实验生成系统，E-Prime 在很大程度上满足了上述要求。

1. E-Prime 的系统组成与功能特点

E-Prime(Experimenter's Prime，简称为 E-Prime)，由美国 PST 公司(Psy-

chology Software Tools, Inc.)开发,是实现行为研究计算机化的一个跨平台系统,其使用原理类似于 Visual Basic 中的 E-Basic 语言,是一个涵盖从实验生成到毫秒精度数据收集与初步分析的图形界面的应用软件系统。E-Prime 是一个可视化的编程语言系统,其核心的功能模块包括 E-Studio、E-Basic、E-Run、E-Merge、E-DataAid、E-Recovery。

(1) E-Studio

E-Studio 是一个图视化的编辑界面,主要用于将 Toolbox 中的编辑对象(E-Object)拖放到 SessionProc 中的程序时间控制线上。一个实验可以由很多个 E-Object 构成,每个 E-Object 都有一组特殊的、可以根据研究需要由研究者自己设定的属性,这些属性将决定对象的行为。譬如,研究者可以设置某个视觉刺激的大小、颜色、在屏幕上呈现的位置、呈现时间(单位是毫秒)、反应方式(键盘、鼠标或者反应盒)、是否需要记录反应时间等。Toolbox 中常用的 E-Object 主要有 Procedure、Slide TextDisplay 等,其具体功能如表 F2-1 所示。

表 F2-1　ToolBox 中主要的 E-Object 及其功能简介

Procedure	用于控制实验的时间进程
TextDisplay	呈现文本刺激
ImageDsiplay	呈现 bmp 格式的单幅图片刺激
Stide	联合呈现文本、图片和声音,或者呈现多幅图片
List	用于定义实验中的自变量、刺激及其属性,如刺激顺序随机化或平衡
SoundOut	呈现 wav 格式的声音文件
FeedbackDisplay	当被试对实验处理作出反应后,自动给予视觉或听觉的反馈
InLine	包含用 E-Basic 语言编写脚本,将其插入到实验中从而扩展 E-Studio的功能

(2) E-Basic

E-Basic 是 E-Prime 中特定的脚本语言。研究者在 E-Studio 中编制的可视化实验程序,只要运行都会被自动地编译成 E-Basic 脚本,并生成文本文件。如果研究者需要一些特殊的、非标准化的操作,而通过 E-Studio 又无法实现,E-Prime允许使用 E-Basic 语言来扩展 E-Studio 的功能。具体做法是:在 Tool-Box 中选择 InLine(如图 F2-1),然后使用 E-Basic 语言编写一个能满足实验特殊

要求的脚本，最后将 InLine 插入到 SessionProc 中，这样当实验运行时 E-Prime 系统会对该脚本进行自动转译。

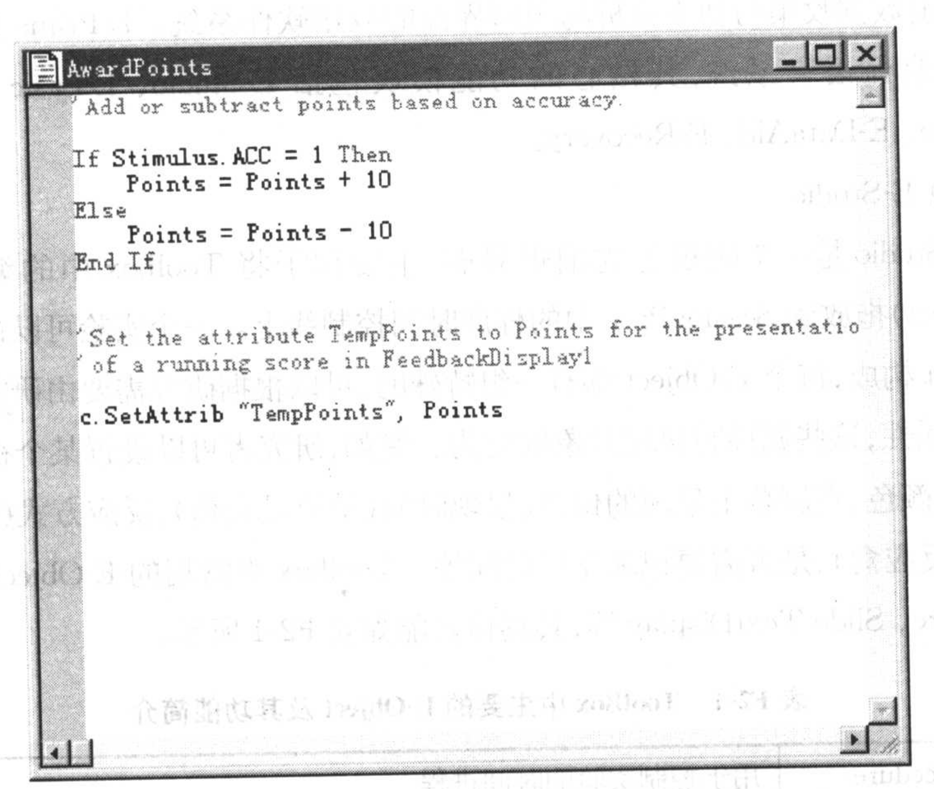


图 F2-1　在 E-Prime 的 InLine 中编写 E-Basic 语言的样例

(3) E-Run

当研究者在 E-Studio 中完全正确地编制好实验后，点击菜单栏中的 E-Run 按钮，系统会自动生成一个关于整个实验的、用 E-Basic 语言编写的脚本。由于整个实验程序被激活，进入运行状态，系统提供毫秒精度的刺激呈现和同步收集被试的反应数据。

(4) E-Merge

E-Merge 是 E-Prime 系统中一个关于数据处理的扩展功能。通过使用 E-Merge，研究者可以把多个数据文件快速、简单地合并，形成一个独立的数据文件，避免每次只能打开一个数据文件的情况。

(5) E-DataAid

E-DataAid 在 E-Prime 中十分有用，具有强大的数据处理功能，它可以过滤数据，譬如研究者只需要被试正确反应的反应时间，那么在 E-DataAid 中选择 Tools 菜单，然后选择其中的 Filter，设定 stimuli. ACC＝1，就能得到被试正确反应条件下的各项数据。另外，E-DataAid 也可以用来编辑数据，并通过导

出系统将数据导入到指定的电子表格或 SPSS 文件中，以便进行更进一步的数据分析。

(6) E-Recovery

在实验程序运行过程中，如果研究者或被试意外终止实验程序，可能会造成数据文件丢失或被破坏，这时就可以用 E-Recovery 恢复文件，把不完整的E-Run 文本文件转换成 E-Prime 的数据文件。

另外，PST 公司还专门开发了与 E-Prime 配套的反应盒 SRBOX 和自控灵敏度的话筒，如图 F2-2 所示。SRBOX 有 5 个反应键，分别对应于 5 种颜色的小灯泡(红、白、蓝、绿、黄)。相对于普通键盘和鼠标来说，SRBOX 的响应延迟时间很短，约为 1.25 ms。除了按键和鼠标反应外，E-Prime 系统还支持用自控灵敏度的话筒进行言语反应以及脚键反应。

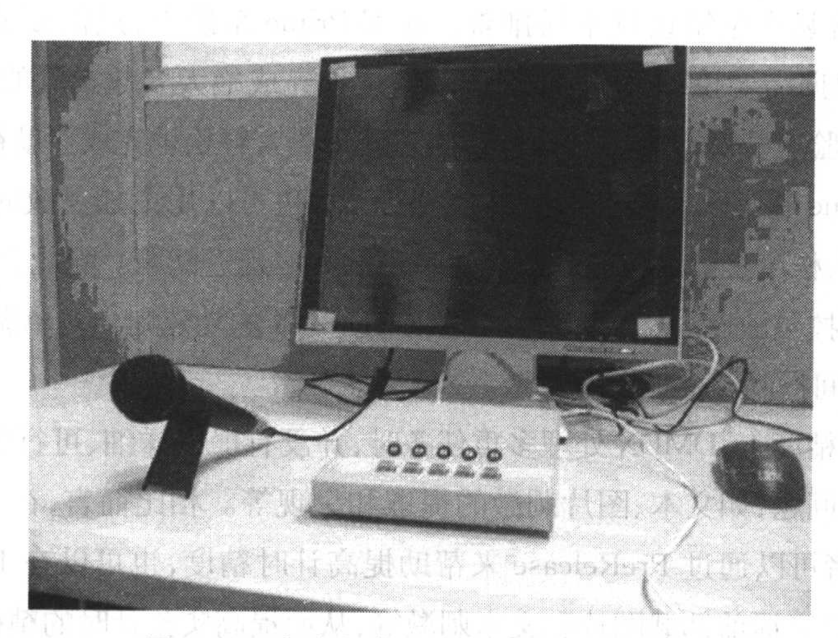


图 F2-2 南京师范大学"E-Prime 实验室"

综合来看，E-Prime 系统具有以下几方面的显著特征:(1)图形化界面的编程环境:研究者可以通过所见即所得的方式选择、拖放实验所需对象，并设定其属性，使得心理实验编程简单化;(2)时间精度非常高:E-Prime 系统提供了 Pre-Release 功能，提高了刺激呈现的计时精度;(3)面向对象的简易 E-Basic 语言:系统提供了许多针对行为研究的增强命令，为编程提供了灵活性，可以帮助实现更加灵活、全面的实验范式;(4)系统提供了一定的扩展空间，如数据分析和导出系统。通过 E-DataAid，研究者可以在 E-Prime 系统内完成初步的数据筛选、整理和分析。即使数据遭到破坏，也可以通过 E-Recovery 进行恢复。

2. 与 DMDX 等实验软件的比较

在引进 E-Prime 以前,国内心理学研究者比较熟悉和经常使用的实验软件是 DMDX。通过以下几个方面的比较,我们可以清晰发现二者的优劣。

(1) 实验设计的编程界面

DMDX 编程界面的可视化程度低于 E-Prime,其刺激控制主要由 RTF 脚本控制,程序结构被分割。而 E-Prime 的 E-Studio 能够直观地呈现整个实验程序的结构。研究者可通过 SessionProc 中的程序时间控制线宏观把握整个实验中事件的流程,从而加强研究者对实验的控制和理解。

(2) 实验结构的搭建

DMDX 有特定的程序语言,研究者需要按照其固有格式编写命令来构建所需的实验程序。如果实验结构较为复杂,研究者编写实验程序将花费较多的时间,而且很容易产生错误且不易排查。在 E-Prime 系统中,只需要简单地拖放 Object,并根据需要设置其属性,就可以快速地完成绝大多数实验的编制;如果研究者对实验有特殊要求,也可以借助 E-Basic 语言轻松地完成。就被试的反应来说,E-Prime 既可以用 PC 键盘和鼠标来完成,也可以用更加灵敏的 SRBox 反应,因此很大程度上解决了计算机硬件带来的响应延迟问题。除此之外,研究者还可以用自控灵敏度的话筒进行声音反应,DMDX 的功能则相对单调。

(3) 时间控制

在计时精度上,DMDX 处理多重任务时,并没有提供详细、可行的方法来改善响应延迟问题,如文本、图片刺激的提取和呈现等。相比而言,在 E-Prime 系统中,研究者可以通过 PreRelease 来帮助提高计时精度,也可以在 InLine 中编写 E-Basic 语言提前卸载图片和文本刺激等,从而提高实验计时的精确性。

(4) 数据的收集和处理

面对大量的原始数据,DMDX 没有提供可扩展的功能模块,无法在系统内完成数据的筛选和初步加工,必须要借助其他软件,如 EXCEL 等完成。在实验控制上,DMDX 很难控制刺激的复杂组合。相比之下,E-Prime 的灵活性和扩展性更强,通过不同的 BlockList、Trial 的组合。如前所述,E-Prime 系统可以在系统内完成数据的初步处理,也可以将数据导入其他分析软件,如 EXCEL、SPSS 等进行分析。DMDX 只能提供简单的反应时间和对错反应,而 E-Prime 的 Logging 提供了 22 种指标,研究者可以根据实验需要任意选择一种加以记录,同时系统会自动以“0”(错误)和“1”(正确)的方式标记正误反应。

(5) 技术支持

由于是商业软件，PST 公司对 E-Prime 系统有专业的技术支持。一方面在 PST 公司的网页上有专门的 E-Prime 板块，有关于该系统的初步介绍、不同 E-Prime版本的性能、购买方法和价格的介绍等，定期提供关于改善系统的免费下载补丁、一些在安装和使用过程中常见的问题及其解决办法；另一方面，PST 公司开辟了专门的 E-Prime Web Support Users，借助该服务，客户可以和公司的专业技术人员交流技术难题，并能得到快速回复。由于是免费软件，DMDX 在技术支持上显得逊色许多。

另外，常用的还有 Presentation 实验设计软件，虽然在国内的使用率不比 DMDX 高，但它和 E-Prime 一样也都是专业的实验编辑软件，可以提供技术支持，但仍然需要写入相当多的程序语句，生成 Sce、Pcl 文件，因此编制复杂实验会耗费研究者大量的时间。

3. E-Prime 在心理学实验教学中的应用

与传统的编程语言相比，E-Prime 实验系统容易掌握、使用方便、计时精度高、记录指标丰富、可以快速生成相对复杂实验的控制程序，并在实验之后进行初步的数据处理，极大地节省了研究者在实验刺激编程上的时间。张智勇等将用于心理实验的编程软件或硬件统称为心理实验系统(Psychological Experimental System，简称 PES)。

心理实验教学是理论联系实际的桥梁，也是提高心理学科学生专业技能和培养学生创新精神、创新思维和创新能力的过程。在心理学教学实验中使用的仪器设备种类繁多，如果按在教学中的用途来划分，其包括课堂演示用和学生实验用两大类。国内现有大多数心理学实验室，尚以传统的心理实验教学为主，显得内容较为陈旧、单调，演示性、验证性实验项目偏多，应用性、设计性、综合性和创造性实验项目偏少，而且学生基本都是按照实验指导书上规定的程序按部就班地完成实验，在整个实验流程中处于较为被动的地位。有研究者曾作过实验心理学教学的调查，发现学生通过一个学期的心理实验课程学习后，离开了实验教师的指导仍不能独立进行实验的设计和操作，更谈不上运用实验方法去研究或解决实际问题。

总之，教学演示实验给学生提供的科学视角存在偏差，如果学生有真实验的体验，就会对科学及其研究过程有更好的理解。由于传统的心理实验教学存在诸多弊病，因此很有必要把科学技术的新成果引入心理实验教学，通过心理学实

验培养学生的动手能力、创造性的思维方法，探索未知和勇于创新的精神。

鉴于 E-Prime 具有的优点，用其作为实验教学系统，可以避免传统实验教学系统的诸多缺点，从而更好地实现心理学实验教学的目标。E-Prime 作为开放的实验教学系统，允许教师和学生以合理的方式任意控制或修改实验中的各种变量和参数设置。由于 E-Prime 实验系统编程的可视化，学生可以直观地“看到”实验设计的整个过程，从而加深对给定演示实验的理解。另外，E-Prime 系统编制实验方法简单易学，学生能够在较短的时间内掌握，从而激发其设计一项研究实验并编制实验程序的热情，有利于学生创造力的发展。

Macwhinney 等(Carnegie Mellon University)在 2001 年启动了一个项目叫 STEP(System for the Teaching of Experimental Psychology)。在该项目中，研究者提供了各个心理学领域下应用 E-Prime 编制的教学材料和许多经典实验，涉及的领域包括感知觉、记忆、注意、生理心理学、心理语言学、发展心理学、认知神经科学等等。通过对这些实验程序和实验设计思想的考察，有助于学生拓宽视野，加强对现代认知心理学及心理学前沿学科的认识。

4. E-Prime 在心理学实验研究中的应用

在国外，E-Prime 已经被广泛地应用于心理学实验研究，成为必备的研究工具。国内，中国科学院心理研究所于 2000 年率先引入 E-Prime 心理学实验系统，成立了 E-Prime 学习交流的平台。随着 E-Prime 系统被国内越来越多的研究人员所熟悉，全国很多心理学研究机构或大专院校都购买了单机版或多机版的 E-Prime 系统。如北京师范大学心理学院、南京师范大学认知科学实验室、中山大学心理系、华南师范大学心理系等，涉及的版本主要是 E-Prime v1. x 系列。据 PST 公司的介绍，最新版本 E-Prime v2. 0 将于 2006 年秋季发布，所有 E-Prime v1. x 的正版用户都可以免费升级。

由于 E-Prime 系统的优良特点，国内研究者和一些心理学硕士生、博士生已经应用该软件完成了许多实验研究，有些成果发表在《心理学报》、《心理科学》等权威学术刊物上，如莫雷等完成的《熟练中英双语者跨语言长时重复启动效应》研究、周治金等完成的《汉语同音歧义词歧义消解的过程及其抑制机制》的研究、胡平等完成的《线索对基本特征刺激加工作用机制研究》等。在视觉认知实验中，研究者对实验的计时精度要求很高，使用 E-Prime 系统能较好地满足其要求。E-Prime 可通过 PreRelease，或者用 E-Basic 语言提前卸载图片，而且 PST 公司还为该系统准备了专门的 SRBOX，有助于研究人员实现毫秒级的

时间精度。

在现代认知神经科学实验中,如 ERP、fMRI 等,研究设备通常需要与外部电脑进行信息通讯。譬如,中国科学院心理研究所就使用 E-Prime,而非 Stim 来呈现实验刺激,完成 ERP 实验。就 ERP 而言,与 fMRI 相比,它以时间的准确性见长,因此实验中刺激呈现时间的精确性十分重要。采用 E-Prime,可以较好地实现刺激事件与脑电记录设备的同步。E-Prime 提供了 WritePort、ReadPort 命令,通过串口或并口与外部设备通讯,使二者尽可能达到同步,其常用命令如下所示:

```
ObjectName.OnsetSignalEnabled = True
ObjectName.OnsetSignalPort = &H378
ObjectName.OnsetSignalData = 1
```

另外,E-Prime 软件可以被应用于眼动实验刺激材料的呈现控制。眼动研究是现代心理学研究中十分活跃的领域,也是近几年来国内心理学科发展速度最快的研究领域之一,已经产出了很多研究成果。就同步刺激切换和数据记录而言,有些眼动仪设备系统会配置厂家研发的刺激呈现软件,从而实现被试电脑和主试电脑刺激的同步切换以及数据的同步记录。但这些刺激软件有的是用高级专业程序编制的,增加了心理学研究者的工作难度;有的眼动仪厂家并不是刺激编辑软件的专家,因此软件的时间精度比较差,难以符合眼动仪设备高时间精度的要求。实际上,E-Prime 系统提供了用数据通讯端口如 COM1,在 InLine 中

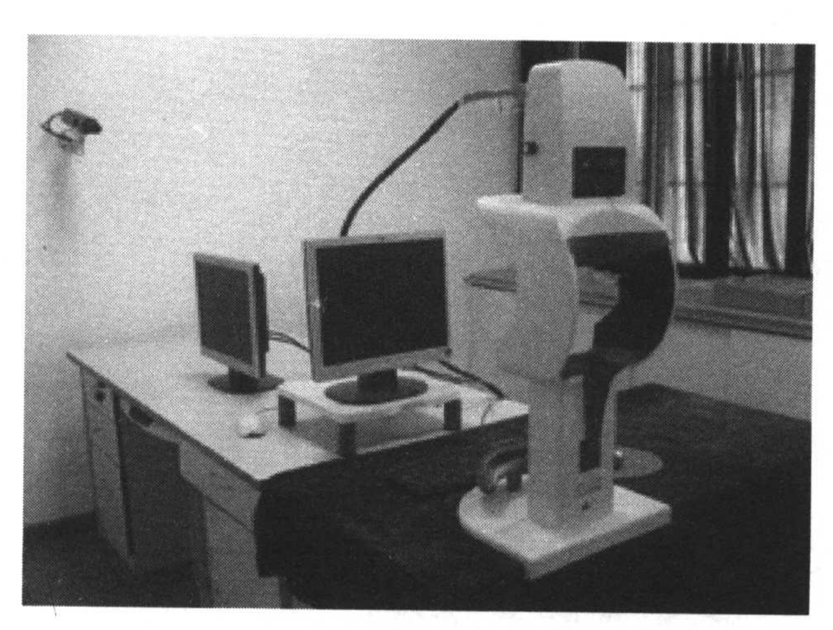


图 F2-3　南京师范大学 SMI 高速眼动仪实验室

编写命令来实现信息共享的方法。通过这种方式，研究者可以轻松地完成主试和被试电脑刺激的同步切换和数据记录。南京师范大学眼动仪实验室使用的SMI的Iview X Hi-Speed高速眼动仪已经实现了使用E-Prime系统编辑与呈现刺激的功能（如图F2-3），通过数据通讯端口实现刺激的同步切换和数据的同步记录，效果良好。

随着心理学实验中计算机技术的进一步发展，E-Prime的应用必然会有效地推动心理学实验教学和实验研究的发展。E-Prime系统简单易学、时间精度高，符合现代认知心理学研究的趋势，而且专门为心理学实验做了优化，在很大程度上减轻了研究者编制实验刺激控制程序的负荷。国内有关心理学研究所和大专院校已经建立或开始着手筹建专门的E-Prime实验室，因此可以预见未来很短的时间内将会出现一大批使用E-Prime实验系统的研究者，同时也会出现更多借助该系统完成的实验研究报告。

附录 3

常用统计检验表

常用统计检验表

1. 随机数字表

Row/Col	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
00000	10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
00001	37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
00002	08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
00003	99019	02529	09376	70715	38311	31165	88676	74397	04436	27659
00004	12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
00005	66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
00006	31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
00007	85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
00008	63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
00010	98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
00011	11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
00012	83452	99634	06288	98083	13746	70078	18475	40610	68711	77817
00013	88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
00015	65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
00016	80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
00017	74350	99817	77402	77214	43236	00210	45421	64237	96286	02655
00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
00019	09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
00020	91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
00021	80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
00023	12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
00024	63606	49329	16505	34484	40219	52563	43651	77082	07207	31790
00025	61196	90446	26457	47774	51924	33729	65394	59593	42582	60527
00026	15474	45266	95270	79953	59367	83848	82396	10118	33211	59466
00027	94557	28573	67897	54387	54622	44431	91190	42592	92927	45973
00028	42481	16213	97344	08721	16868	48767	03071	12059	25701	46670
00029	23523	78317	73208	89837	68935	91416	26252	29663	05522	82562
00030	04493	52494	75246	33824	45862	51025	61962	79335	65337	12472
00031	00549	97654	64051	88159	96119	63896	54692	82391	23287	29529
00032	35963	15307	26898	09354	83351	35462	77974	50024	90103	39333
00033	59808	08391	45427	26842	83609	49700	13021	24892	78565	20106
00034	46058	85236	01390	92286	77281	44077	93910	83647	70617	42941

(续表)

Row/Col	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
00035	32179	00597	87379	25241	05567	07007	86743	17157	85394	11838
00036	69234	61406	20117	45204	15956	60000	18743	92423	97118	96338
00037	19565	41430	01758	75379	40419	21585	66674	36806	84962	85207
00038	45155	14938	19476	07246	43667	94543	59047	90033	20826	69541
00039	94864	31994	36168	10851	34888	81553	01540	35456	05014	51176
00040	98086	24826	45240	28404	44999	08896	39094	73407	35441	31880
00041	33185	16232	41941	50949	89435	48581	88695	41994	37548	73043
00042	80951	00406	96382	70774	20151	23387	25016	25298	94624	61171
00043	79752	49140	71961	28296	69861	02591	74852	20539	00387	59579
00044	18633	32537	98145	06571	31010	24674	05455	61427	77938	91936
00045	74029	43902	77557	32270	97790	17119	52527	58021	80814	51748
00046	54178	45611	80993	37143	05335	12969	56127	19255	36040	90324
00047	11664	49883	52079	84827	59381	71539	09973	33440	88461	23356
00048	48324	77928	31249	64710	02295	36870	32307	57546	15020	09994
00049	69074	94138	87637	91976	35584	04401	10518	21616	01848	76938

2. 正态分布下的 *PZO* 转换表

P	*Z*	*O*	*P*	*Z*	*O*	*P*	*Z*	*O*	*P*	*Z*	*O*
0.01	−2.326	0.0267	0.26	−0.643	0.3245	0.51	0.025	0.3988	0.76	0.706	0.3110
0.02	−2.053	0.0484	0.27	−0.612	0.3308	0.52	0.50	0.3984	0.77	0.738	0.3038
0.03	−1.881	0.0681	0.28	−0.582	0.3368	0.53	0.75	0.3978	0.78	0.772	0.2962
0.04	−1.750	0.0862	0.29	−0.553	0.3424	0.54	0.100	0.3970	0.79	0.806	0.2883
0.05	−1.645	0.1032	0.30	−0.524	0.3478	0.55	0.125	0.3958	0.80	0.841	0.2801
0.06	−1.555	0.1192	0.31	−0.495	0.3529	0.56	0.150	0.3945	0.81	0.877	0.2715
0.07	−1.476	0.1343	0.32	−0.467	0.3577	0.57	0.176	0.3928	0.82	0.915	0.2625
0.08	−1.405	0.1487	0.33	−0.439	0.3623	0.58	0.201	0.3909	0.83	0.954	0.2532
0.09	−1.340	0.1625	0.34	−0.412	0.3665	0.59	0.227	0.3888	0.84	0.994	0.2434
0.10	−1.281	0.1756	0.35	−0.385	0.3705	0.60	0.253	0.3864	0.85	1.036	0.2333
0.11	−1.226	0.1881	0.36	−0.358	0.3742	0.61	0.279	0.3838	0.86	1.080	0.2227
0.12	−1.175	0.2001	0.37	−0.331	0.3777	0.62	0.305	0.3808	0.87	1.126	0.2116
0.13	−1.126	0.2116	0.38	−0.305	0.3808	0.63	0.331	0.3777	0.88	1.175	0.2001
0.14	−1.080	0.2227	0.39	−0.279	0.3838	0.64	0.358	0.3742	0.89	1.226	0.1881
0.15	−1.036	0.2333	0.40	−0.253	0.3864	0.65	0.385	0.3705	0.90	1.281	0.1756
0.16	−0.994	0.2434	0.41	−0.227	0.3888	0.66	0.412	0.3665	0.91	1.340	0.1625
0.17	−0.954	0.2532	0.42	−0.201	0.3909	0.67	0.439	0.3623	0.92	1.405	0.1487
0.18	−0.915	0.2625	0.43	−0.176	0.3928	0.68	0.467	0.3577	0.93	1.476	0.1343
0.19	−0.877	0.2715	0.44	−0.150	0.3945	0.69	0.495	0.3529	0.94	1.555	0.1192
0.20	−0.841	0.2801	0.45	−0.125	0.3958	0.70	0.524	0.3478	0.95	1.645	0.1032
0.21	−0.806	0.2883	0.46	−0.100	0.3970	0.71	0.553	0.3424	0.96	1.750	0.0862
0.22	−0.772	0.2962	0.47	−0.075	0.3978	0.72	0.582	0.3368	0.97	1.881	0.0681
0.23	−0.738	0.3038	0.48	−0.050	0.3984	0.73	0.612	0.3308	0.98	2.053	0.0484
0.24	−0.706	0.3110	0.49	−0.025	0.3988	0.74	0.643	0.3245	0.99	2.326	0.0267
0.25	−0.674	0.3179	0.50	−0.000	0.3989	0.75	0.674	0.3179			

附录 3

常用统计检验表

3. t 分布检验表

(单侧检验用)	0.25	0.10	0.05	0.025	0.01	0.005
(双侧检验用)						
df	0.50	0.20	0.10	0.05	0.002	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

4. *F* 分布检验表

分母自由度	分子自由度														
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248
	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6082	6106	6142	6169	6208
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44
	98.49	99.00	99.17	99.25	99.30	99.33	99.34	99.36	99.38	99.40	99.41	99.42	99.43	99.44	99.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66
	34.12	30.92	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.92	26.83	26.69
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80
	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37	14.24	14.15	14.02
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56
	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	10.15	10.05	9.96	9.89	9.77	9.68	9.55
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87
	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.60	7.52	7.39
7	5.59	4.47	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44
	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47	6.35	6.27	6.15
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15
	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67	5.56	5.48	5.36
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93
	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11	5.00	4.92	4.80
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77
	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71	4.60	4.52	4.41

(续表)

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65
	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40	4.29	4.21	4.10
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54
	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16	4.05	3.98	3.86
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.85	3.78	3.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88

（续表）

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93	2.83	2.74	2.63
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84	2.74	2.66	2.55
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91
	7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89
	7.44	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47

（续表）

1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66	2.56	2.49	2.37
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.46	2.35
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78
	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56	2.46	2.39	2.26
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.43	2.35	2.23
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.40	2.32	2.20

（续表）

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47	2.37	2.30	2.18
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72
	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45	2.35	2.28	2.15
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41	2.32	2.24	2.11
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68
	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36	2.26	2.19	2.06
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33	2.23	2.15	2.03
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64
	6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30	2.20	2.12	2.00
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62
	6.76	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28	2.17	2.09	1.97
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60
	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23	2.12	2.04	1.92
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58
	6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.26	2.20	2.09	2.01	1.89
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57
	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18	2.07	1.99	1.87

5. χ^2 分布检验表

df	0.10	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.96
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.68	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.64
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67
40	51.81	55.76	59.34	63.69	66.77
50	63.17	67.50	71.42	76.15	79.49
60	74.40	79.08	83.30	88.38	91.95
70	85.53	90.53	95.02	100.42	104.22
80	96.58	101.88	106.63	112.33	116.32
90	107.56	113.14	118.14	124.12	128.30
100	118.50	124.34	129.56	135.81	140.17

术语表

1. **α水平**(alpha level) 也叫做显著性水平(level of significance),是假设检验中的统计学显著性标准,即假定只按随机事件而得到相应研究结果的最大允许概率。
2. **暗适应**(dark adaptation) 个体从光亮环境进入黑暗环境,眼睛的感受性或视敏度逐渐提高的过程。
3. **暗适应曲线**(dark-adaptation curve) 描述暗适应过程中视觉感觉阈限随时间变化的曲线。
4. **安慰剂**(placebo) 作为处理或药物的替代品,它没有实际的处理成分。
5. **安慰剂控制组**(placebo control group) 在研究中接受安慰剂作用而不是实验处理的被试组或条件。
6. **安慰剂效应**(placebo effect) 误认为服用了有效药物而产生的心理疗效。在心理实验中,控制组被试也会误认为接受了某种实验处理,由此产生某种心理效应。
7. **按比例分层随机抽样**(proportionate stratified random sampling) 随机抽样技术的一种,它需要先确认总体中包含的子群和各子群在总体中所占比例,然后从每一子群中随机抽取一定量的个体,并使各子群抽取数在样本中所占比例与其在总体中所占比例相同。
8. **白噪声**(white noise) 由各种振幅接近、频率不同的声波复合而成的噪声,其中的各种声波成分不具有谐波关系,在相位上也无系统关系。
9. **保持量**(retention value) 被试对学习材料保持的百分比。
10. **被试**(subject/participant) 研究者或主试安排参与研究的动物或人(许多情况下,我们习惯于把参与实验的人或动物通称为被试)。
11. **被试部分**(participant subsection/subject subsection) 研究报告中“方法”中的一部分,它描述或说明参加研究的那些人或动物。
12. **被试间实验设计**(between-subjects experimental design) 也叫做独立测量设计(independent-measures design)、独立组设计(independent groups de-

sign)或组间设计(between-groups design),是将互相独立但又相等的被试组分派在不同实验处理下进行测量,然后对各组测量结果进行比较的一种实验设计。

13. **被试内实验设计**(within-subjects experimental design) 也称为重复测量设计(repeated-measures design)或组内设计(within-group design),一个被试组参加全部实验处理的实验设计方法。
14. **被试缺损**(participant attrition) 在一项研究进行期间的被试流失,这可能会对研究的内部效度造成损害。
15. **被试选择偏差**(subject selection bias) 基于便利条件而非无偏程序选择被试形成的样本偏差,这是一种影响研究外部效度的因素。
16. **比例量表**(ratio scale) 测量量表的一种,在该量表上,各个类别按顺序排列,而且该量表具有相等单位与绝对零点。
17. **变异源**(source of variation) 进行一系列观测得到一批数据,引起数据变化的原因就叫变异源。
18. **边缘思维论**(marginal thinking theory) 行为主义的代表人物华生认为思维是作为一种整体的躯体的机能。
19. **便利抽样**(convenience sampling) 也称为随意抽样(haphazard sampling),非概率性抽样方法的一种,它对被试的选择是基于可能性或被试是否乐于参加实验,也就是说,因为他们容易得到。
20. **变量**(variable) 可以在数量上或性质上发生变化的事物的属性。
21. **变异**(variability) 对一个分布中分数之间差异量的测量。
22. **辨别反应时间**(identification reaction time) 只有一个反应,它仅与多个刺激中的某一刺激相对应。只有当对应的这个刺激出现时,被试才作出规定的反应,这时测得的反应时间叫做辨别反应时间。
23. **辨别力指数**(index of discriminability) 根据信号检测实验的结果计算出来的能够反映被试信号分辨能力的参数。
24. **标准差**(standard deviation) 方差的平方根,它通过计算与平均数的平均距离来度量数据的差异性。
25. **标准误**(standard error) 对样本统计量和相应总体参数的标准距离或平均距离的度量。
26. **不等组设计**(nonequivalent group design) 准实验设计的一种,在这种设计

中,不同的被试组是在不同条件下形成的,不允许研究者对被试分组进行操纵,因此各被试组被看成是不相等的组。

27. **不随意学习**(incidental learning) 克雷克为证明其记忆的加工水平说而提出的一种实验程序,即向被试提出不同的操作以引导被试对材料进行不同水平的加工,最后出乎预料地要求被试回忆或再认刺激材料。

28. **不应答偏差**(nonresponse bias) 在通讯调查中,有相当数量的被调查者未予回复,使得那些给予回复的个体常常并不是受调查被试全体的代表。这种不应答偏差会影响调查的外部效度。

29. **部分报告法**(partial report procedure) 向被试呈现一系列刺激材料,立即要求其按照某种提示线索报告某一部分的信息,以此缩短报告时间。

30. **残差**(residual) 在方差分析过程中,按照线性模型对因变量观测值的变异量进行分解所剩余部分。残差可看作是不能解释的随机变量引起的变异量,故被称作误差项,用于计算检验统计量 F 时的分母项。

31. **参考文献部分**(reference section) 研究报告的参考文献部分要列举报告中所有引用文献的完整信息,一般是按第一作者姓氏的字母顺序编排。

32. **参与式观察**(participant observation) 观察法的一种,研究者为了便于观察和记录被试的行为,就作为被观察者的一员参与到他们的活动中。

33. **残词补全任务**(word fragment completion task) 让被试填充残缺单词的缺失字母的一种内隐记忆测验。

34. **操纵核查**(manipulation check) 在一项研究中,用来评估被试是如何认识或解释操纵,以及/或评估操纵的直接效应的附加测量。

35. **操作定义**(operational definition) 用于界定或测量一个构念的程序。一个操作定义会详细说明测量一个外部可观察的行为的程序(一系列的操作),并使用该程序所导致的结果作为一假设结构的定义或测量。

36. **测量程序的效度**(validity of measurement procedure) 测量程序对它声称要测量变量的测量程度。

37. **测量过敏**(assessment sensitization) 参见:过敏(sensitization)。

38. **测量量表**(scale of measurement) 一个用来对个体进行分类的类别集合,它具有称名量表、顺序量表、等距量表和比例量表四种类型。

39. **测验效应**(testing effect) 由重复测量经验对被试的影响导致的累积误差。

40. **策划性观察**(contrived observation) 也称为结构性观察(structured obser-

vation),在特意布置的背景或情境中进行的观察。

41. **产生式系统**(production system)　是指计算机和人所能执行的一组活动。只要有一定的条件就能产生一定的活动,共通用表达式就是"If... then..."。

42. **长时记忆**(long-term memory)　经过最初的知觉后已经从意识中消失了的记忆的提取。

43. **尝试-错误说**(trial-and-error theory)　也叫做试误说,是美国心理学家桑代克建构的一种学习理论。他认为人和动物都是在不断尝试和不断错误中解决问题的,因为人和动物可以在这一过程中将成功反应与紧张解除联结起来。

44. **抄袭**(plagiarism)　将别人的思想或语言作为自己的来呈现。抄袭是不道德的。

45. **称名量表**(nominal scale)　测量量表的一种。在这种量表上的各个类别代表被测量变量的定性差异,各类别具有不同的名称但彼此之间不按照系统的方式相互关联。

46. **成熟**(maturation)　指在研究进程中能够影响到被试测量分数的所有生理的或心理的变化对研究内部效度的损害。

47. **成熟差异**(differential maturation)　在一项研究中,成熟因素在一个组引起的效应不同于在另一个组引起的效应,这会造成研究内部效度的下降。

48. **程序部分**(procedure subsection)　研究报告的"方法"中的一个小部分,它描述完成这一研究的每一个步骤。

49. **重复测量 *t* 检验**(repeated-measures t test)　在被试内或匹配实验设计中,对来自同一被试组或两个匹配被试组的两个系列分数平均数差异的统计检验方法。

50. **重复测量设计**(repeated-measures design)　参见:被试内实验设计(within-subjects experimental design)。

51. **重复再现法**(repeated reproduction)　巴特利特的故事记忆测验法,即让被试学习故事之后,按照自己的记忆讲述故事,而且至少讲述两次,两次间隔的时间中不插入学习。

52. **抽样**(sampling)　也称为抽样程序(sampling procedures)或抽样技术(sampling techniques),选择参加研究的被试的过程。

53. **抽样程序**(sampling procedures)　参见:抽样(sampling)。

54. **抽样技术**(sampling techniques)　参见:抽样(sampling)。

55. **抽样偏差**(sampling bias)　参见:选择偏差(selection bias)。

56. **抽样误差**(sampling error)　自然发生的样本统计量与相应总体参数的差异。

57. **处理条件**(treatment condition)　实验研究中,由操纵变量的具体值规定的情境或条件。一个实验包含的两个或两个以上的处理条件与被操纵变量的不同值对应。

58. **纯音**(pure tone)　单一频率的声波,波形为正弦波。

59. **刺激变量**(stimulus variable)　来自外部刺激情境的在心理实验中对被试的心理或行为可能产生影响的各种条件或因素。

60. **代表性**(representativeness)　样本特征能准确反映总体特征的程度。

61. **代表性样本**(representative sample)　在与研究有关的特征方面,与总体具有同样特征的样本。在统计学意义上,它能代表总体。

62. **单盲研究**(single-blind research)　在有安慰剂控制组的对照实验中,被试不了解实验处理安排的研究程序。

63. **单因素方差分析**(single-factor analysis of variance/one-way ANOVA)　一种用于评估两个或两个以上独立组平均数差异的假设检验方法,这里的各个组是由同一个变量或因子的独立值规定的。

64. **单因素两组设计**(single-factor two-group design)　也称为两组设计(two-group design),一种用于比较两组被试或两组分数的设计,每个被试组或每一组分数代表一个因子的一个水平。

65. **倒摄抑制**(retroactive interference)　后学习材料对先学习材料的保持与回忆的干扰。

66. **等距量表**(equal interval scale)　测量量表的一种,在该量表上的类别按顺序组织。该量表具有相等单位,但其零点是一个人为规定的相对值。

67. **等响度曲线**(equal-loudness contour)　响度与频率的关系曲线,即以比较音和标准音在响度上等效时的声级变化为频率的函数所作的曲线。

68. **抵消平衡**(counterbalancing)　在被试内实验设计中,为了使顺序效应和与时间有关的影响降到最小而将被试分成若干小组以便各小组以不同的顺序接受一系列实验处理的实验程序,旨在使每一种可能的处理顺序上都有相等的被试参加实验。

69. **地板效应**(floor effect) 全距效应的一种,指测量分数聚集在量表的低端,测量值很少或没有机会再减小。

70. **第三变量问题**(third-variable problem) 两个变量表现出相关,但却可能是二者均受到第三个变量的影响并因此一起发生变化。

71. **调查法**(method of inquiry) 人们获得知识和问题答案的各种方法。

72. **定额抽样**(quota sampling) 一种便利的非概率抽样方法,它需要先确认样本中要包含的具体子群,然后规定各子群要抽取的个体的定额数,进而完成抽样。

73. **动物关怀和利用委员会**(Institutional Animal Care and Use Committee,简称IACUC),是专门考察在研究中对非人类被试采取的处理方法的委员会。

74. **独立测量 *t* 检验**(independent-measures t test) 在被试间实验设计中,评估两个独立被试组间平均数差异显著性的假设检验。

75. **独立测量设计**(independent-measures design) 参见:被试间实验设计(between-subjects experimental design)。

76. **独立性卡方检验**(chi-square test for independence) 一种评估两个或多个被试组之间比例数差异统计显著性的假设检验方法。

77. **短时记忆**(short-term memory) 在知觉后和离开意识觉察前的信息恢复。

78. **对内部效度的损害**(threat to internal validity) 除实验因子外,为研究结果提供任何其他可能解释的因素都会造成研究内部效度的下降。

79. **对偶联合回忆**(paired-associate recall) 一种记忆任务,先呈现单词对(例如松鼠-大象),然后测验时出示配对的第一个单词,让被试回忆第二个单词。

80. **对数定律**(law of logarithmic function) 也叫费希纳定律(Fechner's law),是德国物理学家费希纳提出来的关于心理量与物理量的关系定律,其公式为:$S = k \times \lg R$,即心理量与物理量的对数成正比。

81. **对外部效度的损害**(threat to external validity) 研究本身具有的制约结果推广性的任何特征,都会造成研究外部效度的下降。

82. **对效度的损害**(threat to validity) 会导致对研究过程的质量或研究结果准确性怀疑的研究成分都会造成研究效度的下降。

83. **对应加工迁移理论**(transfer-appropriate procedures approach) 认为记忆中的实验性分离现象,只是由于不同测验任务所要求的加工过程不同造成的,并非是由于受相互独立的两个不同记忆系统控制所致。

84. **顿悟说**(insight theory) 德国心理学家苛勒提出来的问题解决理论。他认为人和动物是依靠突然领悟来发现问题情境中的完形,从而解决问题的,即发现了问题情境中各个事物间的内在联系或完形,就能够解决问题。

85. **多重记忆系统理论**(multiple memory system theory) 将记忆看成是由多个服务于不同对象目标、遵从于不同操作原则的子系统共同作用而实现的一种功能。

86. **多因素重复实验设计**(multi-factor repeated-measure design) 有两个以上的自变量或准自变量以不同水平的结合而形成不同的实验处理,将抽取来的被试作为一个被试组完成所有实验处理下的实验。

87. **多因素完全随机实验设计**(multi-factor randomized experimental design) 有两个以上的自变量或准自变量以不同水平的结合而形成的多个实验处理,将抽取来的被试随机分成多个组,每个组独立地接受一个实验处理下的观测。

88. **额外变量**(extraneous variable) 也叫做控制变量,在一项研究中,自变量和因变量以外的可能对被试的心理或行为产生影响的其他变量。

89. **二次文献**(secondary source) 对他人工作的描写或概述,其撰写人并不是所讨论的研究或观察的直接参加者。

90. **二因素设计**(two-factor design) 研究中包含两个因子的实验设计。

91. **发展性研究方法**(developmental research strategy) 常用于考察一个人随年龄而发生变化的研究方法。

92. **反应变量**(reaction variable/response variable) 也称为因变量,指由于自变量的变化而引起的被试反应或内外变化,在实验中需要观测和记录的变量。

93. **反应定势**(response set) 在等级量表问题上,被试倾向于采用相同的方式对所有或大部分问题做反应。这时就出现了反应定势。

94. **反应时间**(reaction time) 简称 RT,也叫反应潜伏期,就是个体从接受刺激作用到作出相应反应的时间间距。

95. **反作用**(reactivity) 被试因为正参加的研究或知道正接受的测量而对自然行为进行的修改。

96. **方差**(variance) 通过计算各分数离差方的平均值得到的变异测量。

97. **方差分析**(analysis of variance) 因素型实验数据的核心分析方法,即为了

评估自变量或准自变量对观测变量的影响效应，而按照不同的变异源将数据变异进行分解，以考察自变量或准自变量所引起的因变量数据变化的相对程度。

98. **方法部分**(method section) 研究报告的一个部分，对研究实施过程进行相对详细的介绍，以使其他人了解并能复制你的研究过程。一般包括四方面内容：研究被试、仪器设备与材料、研究设计类型、实验操作程序或主要步骤。

99. **方向性问题**(directionality problem) 相关性研究可以建立两个具有共变关系的变量间的相关关系，但是它不能确定这两个变量中何为"因"、何为"果"。

100. **非等控制组设计**(nonequivalent control group design) 准实验设计的一种，在这种研究设计中，研究者不能随机地安排被试而只能使用现有的被试组，一个组在处理条件下完成任务，另一个组在控制条件下完成任务。

101. **非概率抽样**(nonprobability sampling) 抽样方法的一种，在这种抽样中，总体不完全清楚，所以个体被抽中的概率也不清楚。这种方法往往是依据常识或容易度等因素，并试图保持抽样的代表性和避免偏差。

102. **分贝**(decibel, dB) 表示声音物理强度的单位，以两个声音压力比值的对数表示。其计算公式为 $dB = 10 \cdot \log(P_1/P_0)$，式中 dB 代表分贝，P_1 是较强声音的压力，P_0 是参照声音的压力(常取 0.000 2 达因/平方厘米)。

103. **分层随机抽样**(stratified random sampling) 概率抽样技术的一种，它需要首先确定样本中应该包含的子群，然后从每个子群中随机抽取一容量相等的样本，再合并成所需要的样本。

104. **分组偏差**(assignment bias) 导致各组被试特征有明显差异的被试分组过程，就会出现分组偏差。这种偏差会造成研究内部效度的下降。

105. **符兹堡学派**(Würzburg school) 以屈尔佩为主要代表人物的思维研究学派，除对思维进行了一系列实验研究外，主要贡献是提出了无意象思维的观点。

106. **附录部分**(appendix section) 研究报告中详细说明或列举有关的仪器设备、实验材料及相关数据信息的部分。这些信息虽然有用，但如果放在论文的正文中就会破坏正文信息的连续性。

107. **复合音**(complex tone) 由不同频率的多种声波复合而成的复杂声波。

108. **复制**(replication)　使用原研究中使用的基本程序对一个研究过程进行重复,其意在于对原研究的效度进行检验。这一研究复制要么会因为得到与原研究同样的结果而支持原来的研究,要么会因为其很难得到原来研究的结果而对原来的研究产生怀疑。
109. **概率抽样**(probability sampling)　抽样方法的一种,在这种抽样中,总体是知道的,总体中的每一个个案有确定的被选择概率,而且抽样是基于概率而随机进行的。
110. **概念形成**(concept formation)　概念学习的一种方式,个体以直接经验为基础,从大量的同类事物的不同例证中独立发现和掌握事物的共同关键特征的过程。
111. **干扰说**(interference theory)　认为遗忘是由于其他信息的干扰造成的。
112. **感觉记忆系统**(sensory memory system)　即瞬时记忆,是刚刚输入还未经加工的刺激信息的保持系统,它对信息保持的时间很短。
113. **感觉阈限**(sensory threshold)　指刚刚引起感觉的最小刺激强度,或者说是能够有50%的次数能引起感觉、50%次不能引起感觉的刺激强度。
114. **个体差异**(individual differences)　一个被试与另一个被试的不同特征。
115. **工程心理学**(engineering psychology)　心理学与工程技术交叉的学科,它从心理学的观点研究工程技术、制造和使用中人与技术的关系问题。

116. **构造主义心理学**(structuralism psychology)　冯特的学生铁钦纳发展了他的"内省实验法",形成了构造主义心理学。铁钦纳坚决反对美国当时盛行的机能主义和行为主义倾向,认为经验构成意识,意识包括三种元素性状态:一为感觉,属于知觉之元素;二为想象,属于观念之元素;三为情感,属于情绪之元素。心理学研究的目的就是了解这些元素如何构成了人的经验。
117. **关键词**(key word)　研究报告中对描述研究范围或领域、研究核心内容、采用的方法起重要作用的词汇,或者其在文中出现的频率较高。关键词也用来搜索文献。
118. **关联效应**(contingent aftereffects)　不同刺激之间相互影响而发生的效应,如颜色和方向或颜色和形状。
119. **观察法**(observational method)　在保证被试完全自然存在的前提下,对其心理和行为不加干预的直接观察,据此描述其心理活动规律的方法。

120. **光亮度**(luminance)　指物体表面的明亮程度,它取决于从物体表面反射出来的光能量,也就是说它取决于物体表面接受到的光照度和物体表面的光反射率两个方面。

121. **光强度**(luminous intensity)　指光源的发光功率,以单位时间、单位立体角内发出的光能量为测量标准,单位为坎德拉(candela,简称 cd)。

122. **光通量**(luminous flus)　单位时间内由光源向某一立体角或某一面积上辐射的光能量,也就是在单位时间内穿过某一立体角或某一面积的光能总量,其单位为流明(lumen,简称 lm)。

123. **光照度**(illuminance)　指单位时间从光源照射到单位面积上的光能量,单位为勒克斯(lux,简称 lx)。对于一个点光源来说,光照度与光源的发光强度成正比,与到光源的距离的平方成反比。

124. **光锥**(light cone)　又称颜色锥体,是表示颜色系列特征的圆锥体,它以中间垂直轴代表明度,水平截面圆周的不同方位代表色调,锥体中代表某一颜色的点到中间轴的距离代表饱和度。

125. **归纳**(induction)　也称为归纳推理(induction reasoning),利用相对较小的具体观察集合作为基础构建较大的可能观察集合的表述。

126. **归纳结构问题**(problem of inducing structure)　给出几个条件,而问题解决者必须发现隐含在条件中的结构模式才能解决问题。

127. **归纳推理**(inductive reasoning)　参见:归纳(induction)。

128. **过敏**(sensitization)　也称为测量过敏(assessment sensitization),伴随测量程序本身而出现的被试改变。过敏会影响到研究的外部效度,因为它使得伴随测量程序的观测结果与被试在现实世界中的行为表现不同。

129. **好被试角色**(good subject role)　研究中被试采取的一种迎合研究者假设的反应倾向。

130. **恒定刺激法**(constant stimulus method)　也叫次数法(frequency method)、正误法(true-false method),其特点是:只使用从能经常感觉到的刺激到经常感觉不到的刺激之间的 5～7 个恒定刺激点,而且各刺激点随机呈现,不像最小变化法那样必须按照某种顺序呈现。

131. **恒定法**(keeping invariablenes method)　有些额外变量无法消除或没有条件消除,则使其在整个实验进程中保持在某一恒定水平,其对所有被试的影响基本一致。

132. **横断研究设计**(cross-sectional research design) 指使用不同的被试组,每组被试代表一个不同的年龄或年龄段,在同一时间测量和比较这些被试组。

133. **回忆法**(recall method) 一种对保持的测量方法,即让被试重现学习过的项目以检查其保持量。

134. **混合实验设计**(mixed experimental design) 一种将两种不同的研究策略结合在一起的因素型研究方法,如将被试间设计和被试内设计结合,或将一个实验因子与一个非实验因子结合。

135. **混淆变量**(confounding variable) 指那些不拟研究但却与自变量一起发生系统性改变的变量,它使得被试心理或行为变化的原因变得无法确定。

136. **霍桑效应**(Hawthorne effect) 美国心理学家梅奥从1924—1932年间在霍桑公司的电力工厂开展的一系列霍桑实验中发现:在心理实验中,被试由于参加了实验,感到新奇、受重视,激发了工作积极性进而影响实验目的的实现。

137. **击中概率**(probability of hit) 在信号检测实验中,在有信号条件下,被试作"有信号"的反应的概率。

138. **机体变量**(organism variable) 指可能对被试的心理或行为发生影响的被试自身的特征或身心状态。

139. **即时回忆**(immediate recall) 即在刺激材料呈现完毕立即要求被试进行回忆的方法。

140. **集中趋势**(central tendency) 用一个分数作为确定一个分布中心的统计量数,它为整组数据提供了一个具有代表性的测量值。

141. **计算机模拟**(computer simulation) 通过编程,使用计算机模拟人的心理活动过程。

142. **假设**(hypothesis) 关于两个变量间关系的预测性表述。

143. **假设检验**(hypothesis test) 一种推断性统计程序,它使用样本数据评估关于总体假设的可信度。一个假设检验可以确定研究结果是否达到统计学上的显著性水平。

144. **假设检验说**(hypothesis-testing theory) 概念形成理论之一,是布鲁纳等人于20世纪50年代提出来的。该理论认为,概念形成是一个不断提出假设和不断检验假设的过程。

145. **加因素实验**(plus factor experiment)　斯腾伯格开创的反应时间实验法，他认为，若两个因素的效应相互制约，则这两个因素只作用于同一个信息加工阶段；若两个因素的效应相互独立，两因素的效应可以相加，则这两个因素各自独立地作用于某一特定的加工阶段。他依此假设研究了短时记忆信息提取的过程。
146. **间接启动**(indirect priming)　学习阶段与测试阶段呈现的刺激项目有关但不相同。
147. **减法反应时间法**(method of minus reaction time)　也称为减法法则(minus rule)，是由唐德斯的实验范式扩展而来的，后来成为信息加工心理学的基本研究方法。其基本原理是：当两个信息加工系列具有包含和被包含关系时，即其中一个信息加工系列除含有另一个信息加工系列的所有过程以外，还存在一个独特的信息加工阶段或过程，这两个加工系列需要的时间差就是这个独特的信息加工阶段或过程所需要的时间。
148. **简单反应时间**(simple reaction time)　只有一个反应对应于一个特定的刺激，当该刺激呈现时，被试就立即作出规定的反应，这时测得的反应时间就是简单反应时间。
149. **简单随机抽样**(simple random sampling)　总体中每个个体被抽中概率相等且独立的抽样技术。
150. **节省法**(savings method)　一种对保持量的测量方法，即对先前学习过的材料重新学习并达到原来的熟练程度，比较前后两次学习所用的遍数或时间，以第二次学习节省的遍数或时间作为第一次学习保持量的度量。
151. **结果与分析部分**(results and analysis section)　研究报告中展示研究资料及其统计分析总体情况的部分。
152. **近因效应**(recency effect)　在系列学习材料表列最后的若干项目，相对于中间项目而言，信息的保持更好些。
153. **开窗实验**(experiment of open window)　反应时间实验方法的一种，运用它，能比较直接地测量出每个信息加工阶段的特征，就好像是对复杂的信息加工过程打开了一扇窗户，可以使人一览无余。汉密尔顿和霍克等人在20世纪70年代和80年代利用开窗实验成功地进行了字母转换信息加工过程的研究。
154. **可见光**(visible light)　正常人眼能够觉察得到的光谱，其波长范围在380～

780 纳米之间。

155. **科萨科夫综合症**(Korsakoff's syndrome) 脑外伤急性期过去之后仍长期存留的记忆障碍。

156. **科学方法**(scientific method) 一种获取知识的方法,它利用观察提出假设,然后通过附加的、系统的观察对假设作经验性检验。通常,新的观察又会导致新的假设,如此循环不止。

157. **可得总体**(accessible population) 指目标人群中容易访问的那一部分可能被试,研究者一般都是从这样的可能被试群中选取样本。

158. **可检验假设**(testable hypothsis) 一种假设,其中所有的变量、事件和个体都是真实的,因此是可以被界定和观测的。

159. **可拒绝假设**(refutable hypothses) 一个有可能被证实为假的假设,即一个允许其结果与预期不同的假设。

160. **控制变量**(controlled variable) 参见:额外变量(extraneous variable)。

161. **控制组**(control group) 在一项研究中不接受处理或只接受安慰剂处理的被试组。

162. **口语报告分析**(protocol analysis) 也称为出声思考(thinking aloud),是思维研究的有效方法,即要求问题解决者在解决问题时,边思考边解题,边报告自己是如何思维的,然后研究者对这些口语报告材料进行深入细致地分析。

163. **拉丁方**(Latin square) 一种 $N \times N$ 矩阵,在这种矩阵中,N 个不同处理条件中的每一种处理在每一行和每一列中都只出现一次。它被用来编排处理条件的顺序以达到部分平衡。

164. **拉丁方实验设计**(latin square experimental design) 实验设计的一种,采用循环法来平衡实验顺序或其他额外变量对实验结果影响的实验设计,其特点是:由 p 个拉丁字母排成的一个 p 行 p 列的方阵,每个字母在每一行或每一列都仅出现一次。

165. **乐音**(musical tone) 波形呈周期性变化的声音。通常指听起来和谐悦耳的纯音和由具有谐波关系的纯音所组成的复合音。

166. **累积误差**(progressive error) 指与时间有关而与特定处理无关的因素引起的被试行为或成绩上的变化。常见的累积误差有练习效应和疲劳效应。

167. **历史**(history) 在被试内实验中发生的对被试测量分数有影响的外部事件。

168. **历史效应**(history effect) 在研究进程中发生的对被试的测量分数有影响的外部事件对内部效度的损害。

169. **历史效应差异**(differential history effect) 在一项研究中,历史因素在一个组引起的效应不同于在另一个组引起的效应,这会造成研究内部效度的下降。

170. **联结理论**(connectionistic theory) 桑代克根据大量的实验结果,认为动物解决问题的过程就是一个不断地尝试和不断地错误的过程。在尝试过程中,逐渐把问题情境、成功反应、行为结果强化联结起来。

171. **练习/疲劳差异**(differential practice/fatigue) 在一项研究中,练习或疲劳因素在一个组引起的效应不同于在另一个组引起的效应,这会造成研究内部效度的下降。

172. **练习效应**(practice effect) 当先前实施的测量程序导致被试对其有了额外的熟练度而使分数提高,这种效应会对研究的内部效度造成损害。

173. **两种记忆说**(dual memory theory) 1965年,沃和诺曼正式提出两种记忆说,即认为在人的长时记忆系统之外还存在着短时记忆系统。

174. **临床实验法**(clinical experiment method) 皮亚杰引入的一种儿童心理学实验方法,具体做法是:研究者在与儿童半自然的交往中向儿童提出一些活动要求,让他们看一些实物或向他们提一些特定的问题,儿童可毫无拘束地回答问题,然后研究者针对儿童的反应继续追问和进行观察。

175. **零响度曲线**(0-loudness curve) 等响度曲线中特殊的一条,该曲线上的每一点代表相应频率的声音的压力级在绝对感觉阈限水平。

176. **漏报概率**(probability of miss) 在信号检测实验中,当有信号出现时,被试却作“无信号”反应的概率。

177. **伦理学**(ethics) 研究人的适当行为。

178. **轮廓**(outline) 物体的外形线,当视野中的明度或色调发生了突变而将视野划分成不同的区域时,就有了轮廓。

179. **罗森塔尔效应**(Rosenthal effect) 参见:实验者效应(experimenter effect)。

180. **逻辑理论机**(Logic Thoery Machine,简称LTM) 纽厄尔和西蒙在1956年提出的证明数理逻辑定理的程序。

181. **马赫带**(Mach band) 一种明度对比现象,是奥地利物理学家马赫发现的因不同区域的亮度的相互作用而产生的明暗边界处的主观明度加强现象。

182. **麦科洛效应**(McCollough effect) 视觉刺激诸如方向、颜色、运动等不同特性之间发生的一组视觉后效现象。如交替注视有橙光背景的垂直栅条和有蓝绿背景的水平栅条较长时间后,同时注意有白光背景的垂直栅条和水平栅条,看到的两种颜色发生了互换。

183. **美国心理学会《出版手册》**(第五版,2001)(Publication Manual of the American Psychological Association) 一种描述行为科学领域通用的研究报告写作风格和结构的手册。

184. **美国心理学会伦理条例**(APA Ethics Code) 美国心理学家从事专业的和科学的工作所依据的公认价值体系。这一条例为心理学家提供了在大多数情形下应遵循的一般原则和决策法则,其基本目标是给予那些与心理学家一起工作的个人或团体以权益保护。

185. **幂函数定律**(power function law) 也叫做幂定律、史蒂文斯定律,史蒂文斯认为,当物理量以几何级数增长时,心理量也按几何级数增长,即心理量是物理量的幂函数。

186. **模拟**(simulation) 在研究中创设能模拟或接近行为发生的自然情境。

187. **模拟研究**(simulation experiment) 也称为仿真研究,就是通过仿真或模拟技术创造出与研究问题的实际情形相同或相似的情境,使身处其中的被试与处于实际系统时的心理状态与行为活动相同或相似。

188. **目标总体**(target population) 参见:总体(population)。

189. **内部效度**(internal validity) 一项研究得到的单一的、清晰的对结果的解释程度。

190. **内省法**(introspection method) 参见:自我观察法(self-observational method)。

191. **内隐记忆**(implicit memory) 不能有意识"回忆"的记忆。

192. **内隐记忆测验**(implicit memory tests) 通过间接测验的方法来检测被试意识不到的记忆效果,常用的方法是残词补全和启动测验等。

193. **纽伦堡条例**(Nuremberg Code) 指导对人类被试进行处理的10条伦理规则。形成于1947年纽伦堡法案的"纽伦堡条例"为制订通用的医学和心理

学研究伦理标准奠定了基础。

194. **排列问题**(problem of arrangement) 给出某一事物构成的各种成分,问题解决者必须以某种合理的方式对这些成分进行排列,才能达到问题的目标状态。

195. **潘弄范围**(Panum's fusional area) 1958 年,潘弄发现当某一刺激在双眼网膜成像的位置虽不对应,但如果相互偏离未超出某一范围的限度时,仍可形成单像,这一范围就被称为潘弄范围。

196. **疲劳效应**(fatigue effect) 由于先前的处理条件或测量程序导致被试的疲劳,使后续测量成绩降低。它会损害研究的内部效度。

197. **匹配法**(matching method) 一种被试分组方法,目的在于使某一特定变量在各组间是平衡的。

198. **匹配设计**(matched-subjects design) 对各组进行比较的一种研究设计,在这种设计中,一个组中的每一个被试都与另一组的一个被试匹配。匹配是为了使研究者考虑到相关变量的不同水平下有匹配的相等被试组。

199. **评价法实验**(experiment of evaluation method) 在有无法实验基础上发展而来的一种信号检测实验方法,即要求被试在作信号的有无判断时,同时给出反应的信心等级,这样就可以获得被试在多个判断标准上的信号检测实验结果。

200. **平均差误法**(method of average error) 心理物理学测定感觉阈限的基本方法之一,基本程序是:要求被试按照递增和递减的方向调整一比较刺激的强度,使其在感觉上与另一标准刺激的感觉强度相等,如此重复很多次,以误差绝对值的平均值作为被试的差别感觉阈限。当标准刺激为零时,该差别感觉阈限也就是绝对感觉阈限。

201. **平均数**(mean) 集中趋势的一种量数,它等于被试测量分数的代数和除以所有测量分数的个数。

202. **期望效应**(expectancy effect) 指在实验系列中对反应改变的期望所造成的被试反应偏差。

203. **欺骗**(deception) 有目的性地对有关研究的信息进行隐瞒或对被试进行误导。欺骗包括被动的和主动的两种形式。

204. **欺诈**(fraud) 研究者有意的采用虚假数据和篡改数据。欺诈是不道德的。

205. **启动**(priming) 通过呈现一个启发人想到相关事件的刺激而引发相关认

知的一种技术。例如,单词"医生"容易使人想到"护士"。

206. **启发式策略**(heuristics strategy) 指个体根据自己已有的知识经验,在问题空间进行粗略搜索来解决问题的策略。

207. **前测过敏**(pretest sensitization) 指当行为测量先于处理进行而使被试对处理的反应受到影响。

208. **区分度**(discrimination) 指因变量随自变量变化而变化的敏感程度,即因变量对自变量变化的敏感程度。

209. **全部报告法**(whole report procedure) 向被试呈现一系列刺激项目,立即要求其尽量报告所接受的全部刺激项目。

210. **全距效应**(range effect) 测量分数聚集在量表的一端,天花板效应和地板效应就是全距效应的两种类型。

211. **人工智能**(artificial intelligence) 通过程序驱动而使计算机等人工设备能够完成人的某些智能活动。

212. **人机系统**(human-machine system) 由人和机器组成的通过人机相互作用实现一定功能的系统。

213. **人类工效学**(ergonomics) 应用科学领域中的一门学科,研究个人受到他所使用的工具、机器及其他装备、具体环境的影响,借以帮助人们更容易、更有效、更舒适和更安全地进行工作。

214. **人体尺寸**(body dimension) 用专门仪器在人体特定部位测点测得的尺寸。

215. **人因工程学**(human-factor engineering) 一门将心理学、生理学、解剖学、管理学以及各种工程学知识与方法相结合,探讨人、工作和生活彼此关系的新领域,它的主要功能在于研究如何设计一个最适合人的生活及工作的新环境。

216. **认知方式**(cognitive style) 又称认知风格,是人们对信息和经验进行组织加工时表现出来的个别差异,是个人在感知、记忆和思维过程中经常采取的、习惯化的态度和风格。

217. **认知神经科学**(cognitive neuroscience) 当前实验心理学的前沿领域,它试图探索伴随各种认知过程而出现的神经运动过程,甚至神经原因。

218. **三色说**(trichromatic hypothesis) 主要由英国物理学家托马斯·扬和德国生理学家赫尔姆霍兹提出的解释人所以能形成各种颜色视觉的理论,故又

称扬-赫三色说。该理论认为,人的视觉系统存在三种不同的视锥细胞,分别对长波带、中波带和短波带的可见光敏感,因此视觉中存在红、绿、蓝三种基色,其他颜色则可由三基色混合而成。

219. **色盲**(color blindness) 又称色觉缺陷,指不能正确形成某些颜色或所有彩色的视觉现象。

220. **闪光融合临界频率**(critical fusion frequency, 简称 CFF) 指个体观察一个闪烁光点从看到闪烁到看不到闪烁的频率分解点。

221. **深度视锐**(depth visual acuity) 在深度知觉阈限测量中,以标准深度或主观深度相等区间下限处形成的视角差作为深度判断能力的测定指标,而且把利用双眼视觉判断深度的能力叫做深度视锐。深度视锐用视角差来测量,二者成反比关系。

222. **生态学方法论**(ecological methodology) 强调将研究对象回归到其所应在的生存状态或生存环境中,然后对其进行有限控制的观测。在心理学中,生态学方法强调把人回归到其所在的文化背景中进行研究。

223. **生态学运动**(ecological movement) 20 世纪 80 年代西方心理学界出现的一种研究倾向,特别是发展心理学研究。该运动强调在自然与社会的生态环境中研究被试心理特点。

224. **声图仪**(sonogram display apparatus) 进行声图测量与分析的专门仪器。声图能从三个维度显示声波的振幅、频率与时间的函数关系,还可以标出不同频率成分的相位关系。

225. **声像记忆**(echoic memory) 听觉的瞬时记忆现象,保持时间 4 秒左右,容量在 5 个左右。

226. **时间序列设计**(time-series design) 指实施处理前和实施处理后都进行一系列的周期性观测。

227. **时间知觉**(time perception) 指个体对直接作用于感觉器官的客观事物的持续性和顺序性的知觉。

228. **实验的仿真性**(experimental realism) 在模拟研究中,实验情境对被模拟情境的心理仿真程度。

229. **实验室**(laboratory) 明确用于科学训练或研究的情境,可以是任何被被试或实验参加者看作假设的情境的房子或空间。

230. **实验室实验法**(laboratory experimental method) 指在专门的心理学实验

室里，利用仪器设备控制实验条件、记录被试的心理或行为变化，以推断被试心理活动规律的研究方法。

231. **实验心理学**(experimental psychology) 研究心理实验的基本理论、基本技术并概括和介绍各心理学分支领域中实验研究成果的科学。

232. **实验研究方法**(experimental research strategy) 也称为操纵研究策略，是试图通过操纵变量，同时测量第二个变量并控制所有其他变量来验证因果关系的研究方法。

233. **实验者偏差**(experimenter bias) 实验者的期望或个人信念对研究结果的影响。

234. **实验者效应**(experimenter effect) 指主试在实验中不知不觉的期望、动机、疲劳、厌倦等心理活动对被试产生的颇为微妙的作用，如罗森塔尔效应就是一种实验者效应。

235. **实验组**(experimental group) 在一项实验中接受处理条件的被试组。

236. **实证法**(method of empiricism) 依靠观察和直接感觉经验获得知识的研究方法。

237. **示波器**(oscillograph) 可以用来记录和显示声波频率和振幅的仪器。

238. **事件相关电位**(event related potential，简称 ERP) 一种内源性诱发电位，由于个体积极参与某项活动，特别是参与认知活动的情况下获得的诱发电位。

239. **视敏度**(visual acuity) 眼睛的分辨能力，包括空间视敏度和时间视敏度两种。空间视敏度是眼睛对刺激物空间细节大小的分辨能力；时间视敏度是眼睛对时间间距的分辨能力，通常用闪光融合临界频率来测量。

240. **视野单像区**(horopter) 如果某一刺激物所在的位置点与两个眼睛节点连线的延长线交于两网膜的对应点上，则该物体形成单像。满足这一条件的各刺激点在空间形成一个通过两节点和注视点的圆周，该圆周就被称为视野单像区。

241. **适应**(habituation) 在行为观察中，被试不断地重复出现在观察的场景中，直至这一场景对他不再是一种新异刺激。

242. **手段-目的分析法**(means-end analysis) 指在解决问题时，首先将问题的目标状态与初始状态进行对照，确定二者的最大差异，将减少乃至消除这一差异作为问题解决的最大目标。为此，必须排除其存在的障碍，如此层

层递推，就会把问题解决的目标分解成小的目标阶梯，而采取一些小的措施和问题解决手段，实现一个一个的小的目标，最终使问题得到解决。

243. **首因效应**(primacy effect)　个体对系列呈现的材料的开始信息的记忆效果优于中间项目的情况。

244. **数据库**(database)　集中于一个主题领域(如心理学)的计算机化的交互式工具，被用于与你的研究有关文献的搜索。

245. **双盲研究**(double-blind research)　一种研究方法，在这种研究中，研究者和被试都不知道可预测的结果。

246. **双眼视差**(binocular disparity)　刺激物在双眼形成的视像差。包括两种情况：一种情况是，同一刺激物由于对于双眼来说存在视角差异而形成了不同的网膜像；另一种情况是，物体在双眼网膜上成像的位置有所不同而形成双像。

247. **双眼视轴的辐合运动**(binocular convergence)　指在看远近不同的物体时，双眼视轴的辐合角度发生相应变化，这种生理运动的信息传递到大脑皮层即可成为深度线索。

248. **水平**(levels)　在一项实验中，那些被选择用来创设或界定处理条件的自变量的不同取值就叫水平。在其他类型的研究中，一个因素的不同取值也叫水平。

249. **顺序量表**(ordinal scale)　测量量表的一种，在该量表上的各个类别具有不同的名称并以先后顺序排列(如，第一，第二，第三)。

250. **顺序效应**(order effect)　在一种处理条件下观测的分数由于参加先前处理条件下的实验[后延效应(carryover effect)]或由于研究进程中获得经验[累积误差(progressive error)]而使之发生的变化。

251. **似然比**(likelihood ratio)　以信号检测实验的结果计算出来的能够反映被试信号判别标准高低的参数。

252. **速度-准确性权衡**(speed-accuracy trade-off)　在反应时间实验过程中，反应速度与反应准确性之间存在互换关系，即追求速度会降低准确性，而提高准确性又会牺牲速度。被试在实验中，会根据内外条件，在速度与准确性之间进行权衡，找到一个平衡点。

253. **算法策略**(arithmetic strategy)　按照某种确定的规则或逻辑顺序搜索问题答案的策略。

254. **算子**(operator) 实现问题状态转换的操作方法。

255. **随机分组**(random assignment) 将被试随机分派到各种不同的处理条件下的一种程序或过程。

256. **随机过程**(random process) 从一系列可能的结果中产生一个结果的程序或过程。每一次的结果都无法预测,而且在每一次,每一种可能的结果都有相等的发生概率。

257. **随机化**(randomization) 利用随机方法避免两变量间的系统性变化关系。它可以破坏任何可能的变量间系统性变化关系。

258. **讨论部分**(discussion section) 研究报告中的一个部分,它主要是对假设的重述、结果的解释和概括,以及对结果的意义和可能的应用进行讨论。

259. **天花板效应**(ceiling effect) 全距效应的一种,测量分数聚集在量表的高分端,使分数提高的可能性很少甚至没有。

260. **条形图**(bar graph) 一种频数分布图,它使用直立的长方条反映类别测量或顺序量表中每个分数的频数。

261. **听觉掩蔽**(auditory masking) 指由于某一声音刺激的存在而使另一声音刺激的强度阈限提高的听觉现象。

262. **通用问题解决者**(General Problem Solver,简称 GPS) 纽厄尔和西蒙在1972 年编制的通用问题解决的产生式系统。

263. **同行评审**(peer review) 这是研究者提交用于发表的研究报告时,许多论文都要经历的编辑程序。在典型的同行评审过程中,杂志社的编辑和同领域的专家对论文进行非常详细的评审。评审者严格细致地审察研究的每一个方面,其基本目的是对研究的质量及其对科学知识的贡献作出评估,也可能对研究或发现中的可疑点进行检测。

264. **图像记忆**(iconic memory) 视觉的瞬时记忆现象,保持时间 300 毫秒左右、容量至少在 9 个项目以上。

265. **图形后效**(figural aftereffects) 指注视一个图形一段时间后,对随后感知别的图形的影响。

266. **图形掩蔽**(figure masking) 在视知觉过程中,对目标刺激物的知觉受到同时或继时出现的其他刺激的影响而变得模糊或发生变异。

267. **外部效度**(external validity) 指能够将研究结果推广到实验情境以外的人、背景、测量和特征的程度。

268. **外显记忆**(explicit memory) 对先前信息的有意识回忆。
269. **外显记忆测验**(explicit memory tests) 需要人们有意识地记住具体事件的记忆测验。
270. **完形主义心理学**(Gestalt psychology) 音译为“格式塔心理学”,其基本观点是:心理现象都是有组织的、不可分割的整体,即是一个完形。人和动物都可以通过内部的心理完形发现外部刺激情境中的完形,于是发生知觉,并解决问题。
271. **文献检索**(literature search) 熟悉你的主题领域中当前正在进行的研究,并找到相对较少的一系列论文的过程。这些论文可以为你提供研究思想,以及为你正要进行的新的研究提供合理性说明和研究基础。
272. **问题**(problem) 当人们不能采取显而易见的直接手段达到目的时,就有了问题,它其实就是人们面临的一种认知情境,在这种情境中人们要达到目标存在某些障碍。
273. **问题空间**(problem space) 问题解决的中介状态以及实现这些状态转换的操作方法,由算子连接问题中介状态构成。
274. **问题行为图**(problem behavior graph) 在认知操作过程中,随着问题解决者采取某种措施,问题的状态就会发生不断变化。把这些状态按照操作的顺序排列出来就构成了一个由各种相互连接的问题状态分布图。
275. **无意象思维**(imageless thinking) 符兹堡学派提出的思维理论,认为思维不能归结为感觉和意象,而是存在一种非感觉的、非意象的思维元素。
276. **无意义音节**(nonsense syllable) 用语音材料或字符组成的刺激串,音素间没有明显的语义联系。艾宾浩斯记忆实验采用的就是无意义音节材料。
277. **误差项方差**(error variance) 随机因素引起的一种变异量,它被用作方差分析中计算 F 比率的分母项。
278. **析因设计**(factorial design) 一种包含两个或更多个因素的研究设计。
279. **习惯效应**(habit effect) 指在一个实验系列中,被试连续采用某种相同的反应方式则会导致一种习惯性偏差,即导致被试的反应改变滞后。
280. **系列位置曲线**(serial position curve) 保持作为信息输入位置函数的曲线。通常,最初的几个项目和最后的几个项目比中间的记忆效果好。
281. **系列效应**(series effect) 一种实验条件下的操作会影响被试在另一种后续实验条件下的操作或反应。

282. **系统抽样**(systemanic sampling) 在一个总体的个体表列中,先随机选择一个起点,然后选择每第 n 个被试以形成一个样本的抽样技术。

283. **先定概率**(prior probability) 即实验之前被试所了解到的信号出现的相对概率。

284. **显著性水平**(level of significance) 参见:α 水平(alpha level)。

285. **现场**(field) 指能被被试感知为自然环境的研究情境。

286. **现场研究**(field study) 在一种被试感觉为自然环境的情境中进行的研究。

287. **线图**(line graph) 用直线连接起来的各点显示不同组或不同处理条件得到的平均数图示。

288. **相等时间取样设计**(equivalent time-samples design) 一种由一个长时间的序列观察组成的准实验时间序列设计,在实验时间序列中交替地施加和撤销一个实验处理。

289. **消除法**(removing method) 控制额外变量的方法之一,即避免有些可能影响被试心理或行为的因素在实验过程中出现。

290. **效度**(validity) 指测量的有效性,即因变量能不能确实测量到它所要测量的被试心理或行为的变化。

291. **心理量表**(psychological scales) 刺激的物理强度与其引起的心理强度的对照表。

292. **心理物理学**(psychophysics) 关于身心之间或外界刺激量与内部活动心理量之间的函数关系或依存关系的严密科学,是介于物理学与心理学之间的一门交叉学科。

293. **心理旋转**(mental rotation) 指在空间知觉加工过程中进行的一种心理上的旋转操作,从而获得正确知觉经验的历程。

294. **《心理学文摘》**(Psychological Abstract) 一装订成册的连续出版物,是搜索心理学文献和查找与你的研究主题相关文章的非计算机化工具,主要包含摘要或概述、主题索引、作者索引等信息。

295. **《心理学文摘》光盘数据库**(PsycINFO) 一种能提供与《心理学文摘》同样信息的计算机化数据库。

296. **心理学资源数据库**(PsycLIT) 一种能提供与《心理学文摘》同样信息的计算机化数据库。

297. **新奇效应**(novelty effect) 当被试在一种研究情境中(一种新异的情境),其感知和反应与其在正常的或真实的世界中不同而对研究的外部效度产生的影响。

298. **信度**(reliability) 指在相同或相近条件下测量结果的稳定性或前后一致性程度。同一个体在相同条件下接受测量,如果测量程序是可信的,那么应该得到相同或相近的测试结果。

299. **信号接收者操作特征曲线**(receiver operating characteristic curve) 简称ROC曲线,是以一系列不同条件下信号检测实验的虚报概率为横坐标、击中概率为纵坐标作出来的二维曲线。

300. **形状**(shape) 轮廓从视野中分割出来的一个面积。

301. **形状后效**(shape after-effect) 又称为弯曲后效(curvature after-effect),是指由于图形方向导致的图形形状知觉的变化。

302. **形状知觉**(shape perception) 空间知觉的一种,是个体对物体轮廓和细节的整体反映。

303. **虚报概率**(probability of false alarm) 在信号检测实验中,没有信号出现时,被试作"有信号"的反应的概率。

304. **序列重建法**(series reconstruction method) 就是先向被试呈现一个刺激序列,给予其一定的学习时间,而且要求被试记住刺激项目间的位置关系,然后将这些刺激的顺序打乱再呈现给被试,让被试根据自己的信息保持恢复刺激材料的原有顺序。

305. **选择反应时间**(choice reaction time) 有多个反应,各自对应于一个指定的刺激。当某一刺激呈现时,被试必须作出与之对应的反应,这时测得的反应时间就叫做选择反应时间。

306. **选择偏差**(selection bias) 也称为抽样偏差,是以增加有偏样本发生概率的方式选择被试时出现的偏差。

307. **选择性探索**(selective explore) 就是选择可能性最少而又获得信息最多的环节开始搜索解决问题的方法和途径。

308. **眩光**(glare) 由于光源或反射面的亮度过大,或由于光源或反射面与背景的亮度反差太大造成的视觉现象,可引起视觉的不舒服感,降低视觉工效。

309. **延迟回忆**(delayed recall) 在刺激材料呈现完毕,插入一项无关任务使回忆延迟一定的时间。

310. **研究报告**(research report) 描述实验研究过程及其成果的书面文件,它要清晰表达的内容主要包括研究目的、引发研究的相关文献背景、实验研究方法、研究结果与分析,以及对研究结果的讨论和解释等。

311. **研究程序**(research procedure) 对具体研究过程逐步地精确描述。

312. **研究方法**(research strategy) 由希望解答的问题性质决定的完成研究的方法。

313. **研究伦理**(research ethics) 关心的是研究者的责任,他必须对那些受其研究过程或研究报告影响的所有人诚实和尊重。研究者通常受到一些伦理规则的制约,这有助于他们作出适当的决策和选择适当的行动。美国心理学会(American Psychological Association,简称 APA)制订了一系列关于心理学研究中的伦理准则(APA, 1992)。

314. **研究设计**(resaerch design) 实施研究的总体计划,设计中要明确采用组内设计还是组间设计、进行单组内比较还是多组间比较,以及研究几个变量等。

315. **眼动**(eye movement) 即眼球的运动,它有三种基本方式:注视(fixation)、眼跳(saccade)和追随运动(pursuit movement)。

316. **眼动仪**(eye tracking apparatus) 记录眼球运动或视线移动轨迹的大型仪器。现代眼动仪的结构一般包括四个系统:光学系统、瞳孔中心坐标提取系统、视景与瞳孔坐标叠加系统、图象与数据的记录与分析系统。

317. **眼跳**(saccade) 头部不动,视线从一个物体或方向转移到另一个物体或方向时眼球的运动。

318. **演绎**(deduction) 也叫做演绎推理,是利用一个一般性的观点作为基础而得到有关具体事实的结论。

319. **样本**(sample) 选自总体的一个个体集合,通常用以在研究中代表总体。

320. **一项研究的效度**(validity of a research study) 研究结果的真实性与准确性。效度是根据研究过程的质量与结果的准确性决定的。

321. **仪器设备与材料**(apparatus and materials) 包含在研究报告方法部分,它一般要说明研究中使用的器材(如设备)或材料(如调查问卷、实验用词单等)。

322. **遗忘曲线**(forgetting curve) 也叫艾宾浩斯曲线(Ebbinghaus curve),是对学习过的材料的保持量随时间而逐渐下降的曲线。

323. **遗忘症**(amnesia) 记忆障碍的一种,一定时间内对生活经历的记忆完全丧失。

324. **因变量**(dependent variable) 在对不同被试组或不同处理条件进行比较的研究中,每个被试都要进行测量的变量。就一项实验来说,因变量要被观测以便确定其是否随着自变量的操纵而变化。

325. **因素(或因子)**(factor) 在一项研究中,能区分一系列实验组或处理条件的变量。实验设计中的一个因素就是一个自变量或准自变量。

326. **因素间的交互效应**(interaction between factors) 指因素型实验设计中,除主效应之外的各处理条件或处理单元间的平均数的差异。也可以说是一个因素的效应依赖另一个因素的不同水平。这种交互作用可以通过一个二因素设计的平均数图示中出现的不平行线(即汇聚或交叉)来显示。

327. **因素型实验**(factorial experiment) 也称函数型实验,它是从艾宾浩斯研究记忆开始的。这种实验的任务不是精确地分析认识过程,而是试图找出一定现象的原因,或者阐明两个变量间的函数关系。

328. **引言部分**(introduction section) 研究报告正文的第一部分,它要表达的是这一研究课题的逻辑发展,包括对相关背景文献的回顾、研究问题或假设的陈述,以及对被用来回答问题或检验假设的方法的简单描述。

329. **引用**(citation) 对研究报告中提到的特定事实或观点的源出版物的作者及发表年代的确认。

330. **有偏样本**(biased sample) 指与来自某一总体的其他代表性样本具有不同特征的样本。

331. **有无法实验**(experiment of yes-no method) 一种信号检测实验方法,在不断重复的实验中,每一次给被试呈现的实验条件,可能存在要其检测的信号,也可能没有信号只有噪声。要求被试每一次判断有无信号,并作出“有”或“无”的反应。实验结束后,根据被试的击中概率和虚报概率计算其信号辨别力指数和似然比。

332. **诱发电位**(evoked potential, EP) 也称诱发反应,指神经系统在接受内外环境刺激信息所产生的特定神经电活动。

333. **元认知**(metacognition) 认知主体对自己认知活动的监控和认识,包括元认知知识、元认知体验和元认知调控三种基本成分。

334. **原始文献**(primary source) 由实际进行研究或观察的人撰写的第一手观

察报告或研究结果报告。

335. **运动视差**(motion parallax) 指在运动过程中形成的对远近不同物体的不同运动速度或方向的知觉印象,它可以作为深度判断的重要线索。

336. **运动知觉**(movement perception) 个体对物体运动方向和运动速度的知觉。

337. **再认法**(recognition) 一种对保持的测量方法,即判断被试对信息熟悉度的一种方法。

338. **噪声**(noise) 由多种纯音复合而成的,但由于这些纯音之间没有谐波关系,所以组成的复合音不再具有周期性,呈现出不规则的波形。

339. **摘要**(abstract) 研究报告组成部分,它要用非常简练的语句表述研究的目的、研究的手段和方法、研究的主要结果或发现。

340. **整群抽样**(cluster sampling) 从一个总体中随机选取组而不是个体的概率抽样技术。

341. **正确否定概率**(probability of correct rejection) 在信号检测实验中,当没有信号出现时,被试作“无信号”反应的概率。

342. **知会同意表**(consent form) 研究者提供的书面声明,上面包含所有知会同意的项目和被试签名行。研究实施之前,要将知会同意表提供给可能被试,以便他们了解在决定是否愿意参加研究之前需要知道的全部信息。

343. **知觉定势**(perceptual set) 个体对某一特定知觉活动表现出来的准备性心理倾向。

344. **知觉图式**(perception schema) 将过去感知过的一些刺激信息有机结合在一起构成的一个结构稳定的内部表征。

345. **直接启动**(direct priming) 也叫做重复启动(repetition priming),指学习阶段呈现的刺激项目与测试阶段呈现的刺激项目完全相同。

346. **指示语**(instruction) 主试向被试交代任务时所讲的话。

347. **指示语效应**(instruction effect) 指示语对被试的心理或行为发生的影响。

348. **志愿者偏差**(volunter bias) 由于志愿者不能完全代表一般总体,进而造成研究的外部效度下降。

349. **制度审核委员会**(Institutional Review Board, IRB) 一个对提出的所有研究计划进行审核的委员会,其职责是确保这些研究在对人类被试进行处理时要保护他们的权益。

350. **中位数**(median) 集中趋势的一种量数,它将分数分布划分为两半,有50%的被试的分数大于这个值,50%的被试的分数小于这个值。

351. **主动欺骗**(active deception) 也称为误导(commission),指有意地向被试呈现关于研究的错误信息,其最常见的形式就是在关于研究目的问题上误导被试。

352. **主观轮廓**(subjective contour) 观察者对轮廓不明确的刺激,凭借个人的知识经验而赋予其一种轮廓的知觉。

353. **主效应**(main effect) 在因素型研究中,一个因素各水平间观测量的平均数的差异。

354. **注视**(fixation) 指将眼睛的中央窝对准某一物体的时间超过 100 毫秒,在此期间被注视的物体成像在中央窝上,获得更充分的加工而形成清晰的像。

355. **转换问题**(problem of transformation) 给出问题的起始状态和目标状态,而且问题结构成分均已知,只是其存在的状态不一样。

356. **追随运动**(pursuit movement) 当被观察物体与眼睛存在相对运动时,为了保证眼睛总是注视这个物体,眼睛会追随物体移动或转动。

357. **准实验研究方法**(quasi-experimental research strategy) 研究方法的一种,它像实验方法一样,也试图降低内部效度受到的影响进而寻求因果关系结论,但是它缺少真实验研究方法中必需的对变量的操纵与控制。准实验研究方法通常是对现成被试组进行比较或使用时间变量界定处理条件。

358. **准自变量**(quasi-independent variable) 指在准实验和非实验研究中,区分被比较的组或处理条件的变量。准自变量类似于真实验中的自变量。

359. **自变量**(independent variable) 在实验中由研究者操纵的变量。在行为研究中,自变量通常具有两个或更多处理条件以作用于被试。

360. **自发电位**(spontaneous EEG) 指在没有特定刺激的情况下,大脑神经系统活动自发产生的电位及其变化。

361. **自然观察**(naturalistic observation) 研究者尽可能不施加干预而在自然情境中对行为进行观察的方法。

362. **自然实验法**(naturalistic experimental method) 在个体自然的生活、学习或工作条件下,由研究者主动改变或创设某种情境条件,有意引起研究对象某种心理活动的发生,从而研究心理现象的一种方法。

363. **自我观察法**(self-observational method)　也叫内省法,是指冯特采用的在严密控制条件下,被试作自我观察的心理学研究方法。

364. **自由度**(degrees of freedom)　概括地说,在样本变异统计分析中独立的变异数叫做自由度。如一个样本的 n 个分数的变异被计算时, $n-1$ 就是其自由度。

365. **自由回忆实验**(free-recall experiments)　将一系列材料全部呈现给被试,要求其学习一定的时间,然后进行自由回忆,即对回忆顺序等不加任何限制。

366. **总体**(population)　也叫目标总体,是研究者感兴趣的个案总体。通常总体虽然不能全部参加研究,但研究的结果却可以推广到总体。

367. **纵向研究设计**(longitudinal research design)　通过跨越一定时间的一系列观测来考察发展的研究设计。典型的形式是在不同的时间点上对同一组被试进行测量。

368. **最小变化法**(minimal-change method)　也称极限法(limit method)或系列探索法,是心理物理学经典方法之一,它的主要特点就是刺激按照“递增”或“递减”的两种系列逐级进行微小的变化,以探测被试对刺激有无觉察的反应变化的转折点,即阈限的位置。

369. **作业疲劳**(fatigue for performance)　当个体作业负荷达到一定量后,其身心紧张度持续下降,反应能力降低。这是机体正常的生理反应,起预防机体过劳的警告作用。

附 录 5

PsyKey 心理学实验教学系统简介[①]

PsyKey 心理教学系统(大学版)是一套为高等院校及师范学校的心理学、教育学、体育及其他相关专业从业人员从事心理学教学和心理学专业研究工作而开发的计算机教学实验系统。该软件可完成 53 个心理学实验的操作或演示,覆盖发展心理、认知心理、实验心理、普通心理等心理学多个分支学科的基础和重点教学内容。

该系统生动的多媒体演示功能让单调的课堂实验教学变得趣味盎然;合理的硬件设计使分组实验准确可靠;强大的数据收集处理功能使实验数据快速生成为最后的科研结果;灵活的实验参数及实验素材设置,便捷的实验列表管理,让实验设计随你而动;网络版更可以让您"掌一机而知全局"。

一、PsyKey 系统的功能

1. 实验管理功能

参数设置:正式实验参数可根据老师实验设计的需要自由设置,大大增加了实验教学的灵活性、研究性和实用性。

个性化实验列表:用户可根据教学要求自由设定实验列表,满足不同的实验教学要求。

结果报表管理:正式实验完成后,系统自动生成包括图表和原始数据在内的结果报表。结果报表可被查询、保存、打印。作为学生实验报告的原始数据记录,方便学生撰写高质量的实验报告。

2. 教学功能

教学资料参考:系统提供了实验的形成和发展、研究发现、实用意义等大量的实验相关背景材料和实验所涉及的一些重要的心理学名词、术语的解释,可作为教师备课和学生写实验报告的参考,也可用于课堂教学演示。

① 详见:北京心灵方舟科技发展有限公司:《PsyKey 心理教学系统实验手册》,2004 年 7 月。

多媒体教学:活泼的互动型演示实验可为多媒体教学课件。

3. 网络功能

网络数据传输:实验结束后,原始实验数据及结果报表将通过网络自动传输至教师管理的服务器,方便教师批阅和统计研究。

二、系统完成的实验内容

PsyKey 心理教学系统(大学版)可完成的实验包含五大类:

1. 心理学演示实验

颜色视觉、知觉的整体性、知觉的选择性、知觉的理解性、知觉的恒常性、深度知觉、错觉、观察力、有意注意与无意注意、机械记忆与意义记忆、表象和想象等。

2. 儿童心理实验

思维策略、辨别学习的策略、空间认知的发展、天平实验、认知方式、儿童道德判断的发展、小学生推理能力的测定。

3. 基本心理能力的实验测定

注意广度、注意分配、时间知觉、速度知觉、短时记忆广度、空间位置记忆广度、工作记忆广度、再认能力测定、简单反应时、选择反应时和辨别反应时。

4. 经典心理实验

明度的差别阈限、音高的差别阈限、线段长度差别阈限、制作颜色爱好顺序量表、部分报告法实验、信号检测实验、系列位置效应、句子理解速度、似动现象、学习迁移、STROOP 效应、河内塔、反应时间作为因变量的优越性。

5. 认知心理实验

概念形成、表象的心理旋转、短时记忆的视觉编码、学习策略(空白试验法)、短时记忆的信息提取、记忆的加工水平、视觉搜索中的非对称性、内隐记忆、记忆错觉、无意识知觉的实验研究(错误再认)、汉语词汇加工过程的抑制机制。

参考文献

中文部分

[美]John B. Best 著,黄希庭主译:《认知心理学》,中国轻工业出版社 2000 年版,第 335—336、390 页。

[美]弗雷德里克·J·格拉维特、罗妮安·B·福泽诺著,邓铸、姜子云、蒋小慧等译:《行为科学研究方法》,陕西师范大学出版社 2005 年版。

[美]墨顿·亨特著,李斯译:《心理学的故事》,海南出版社 1999 年版。

[美]唐纳德. A. 诺曼著,梅琼译:《设计心理学》,中信出版社 2003 年版。

Allen L. Edwards 著,毛正中等译:《心理研究中的实验设计(第五版)》,四川教育出版社 1996 年版,第 381—439 页。

B. H. 坎特威茨、H. L. 罗迪格、D. G. 埃尔姆斯著,郭秀艳等译:《实验心理学》,华东师范大学出版社 2001 年版,第 566—567 页。

Duane, P. Schultz & Schultz, Sydney Ellen 著,时勘等译:《工业与组织心理学》,中国轻工业出版社 2004 年版。

R. 赖丁、S. 雷纳著,庞维国译:《认知风格与学习策略》,华东师范大学出版社 2003 年版。

巴特利特著,黎炜译:《记忆:一个实验的与社会的心理学研究》,浙江教育出版社 1998 年版。

白学军、沈德立:《初学阅读者和熟练阅读者阅读课文时眼动特征的比较研究》,《心理发展与教育》,1995, 11(2):6—12。

北京师范大学等四校联合编写组:《人体解剖生理学》,高等教育出版社 1982 年版,第 111—132 页。

蔡厚德:《刺激的知觉辨认难度与大脑两半球间的分布式加工》,《心理学报》,2005, 37(1):14—18。

蔡启明、余臻、庄长远:《人因工程》,科学出版社 2005 年版,第 206—207 页。

车文博:《西方心理学史》,浙江教育出版社 1998 年版,第 368 页。

车文博:《心理学 150 问》,辽宁人民出版社 1987 年版,第 26—29 页。

陈文锋、崔耀、张建新:《心理实验系统 E-Prime 介绍及其应用》,《心理科学》,2005, 28 (6): 1456—1458。

陈向阳、沈德立:《中小学生阅读寓言过程的眼动研究》,《心理科学》,2004, 27(4):777—780。

陈燕丽、史瑞萍、田宏杰:《阅读成语时最佳注视位置的实验研究》,《心理科学》,2004, 27(2):278—280。

陈英和:《认知发展心理学》,浙江人民出版社 1996 年版,第 312—340 页。

邓铸:《高中生物理问题表征机制的实验研究》,《华人心理学报(香港)》,2004(2): 143—179。

邓铸:《文化分裂及对当代认知研究范型的反思》,《南京师大学报(社会科学版)》,2001(2):91—97。

邓铸:《眼动心理学的理论、技术及应用研究》,《南京师大学报(社会科学版)》,2005, 1:90—95。

邓铸:《元记忆教育训练的认知研究》,《华东师范大学学报(教科版)》,1997, 4:68—73。

邓铸:《知识丰富领域问题表征与解决策略》,《宁波大学学报(教育科学版)》,2002, 24(1):32—36。

邓铸:《认知研究生态学趋向及两种文化的缓和》,见杨鑫辉:《心理学探新论丛》,南京师范大学出版社 2000 年版,第 159—169 页。

董奇:《中小学生阅读能力与元认知的发展与培养》,北京师范大学博士论文,1988 年。

恩格斯:《自然辩证法》,人民出版社 1971 年版。

方富熹、方格、M. 凯勒:《对友谊关系社会认知发展的跨文化比较研究》,《心理学报》,1994, 26(1):44—50。

方富熹、方格、王文忠:《小学低年级儿童对友谊关系的认知发展》,《心理学报》,1993, 25(1):1—7。

方俐洛、虞积生、高晶:《中国正常男青年的深度视觉阈限值的测定(Ⅱ)》,《心理学报》,1980, 12(3):303—306。

傅荣、翟宏:《行为、心理、精神生态学发展研究》,《北京师范大学学报(人文社科版)》,2000(5):110—111。

傅小兰:《探索问题解决的奥秘:表征与策略》,见中国心理学会编:《当代中国心理学》,人民教育出版社 2001 年版,第 37—42 页。

高觉敷:《西方近代心理学史》,人民教育出版社 1982 年版,第 115—137 页。

高明凯、石安石:《语言学概论》,中华书局 1985 年版。
高申春:《冯特心理学遗产的历史重估》,《心理学探新》,2002(1):3—7。
葛列众、朱祖祥:《不同反应方式对双作业操作信息干扰的影响》,《心理科学》,1992, 21(3):226—230。
管键:《生态系统心理治疗的理论述评》,《赣南师范学院学报》,2002(2):80—84。
郭伏、杨学涵:《人因工程学》,东北大学出版社 2001 年版。
郭力平:《学前儿童心理发展研究方法》,上海教育出版社 2002 年版,第 239—243 页。
郭秀艳:《实验心理学》,人民教育出版社 2004 年版,第 459—469 页。
贺革:《试论心理学研究方法的新进展》,《长沙大学学报》,2000(1):66—68 页。
赫葆源、张厚粲、陈舒永:《实验心理学》,北京大学出版社 1983 年版,第 538 页。
胡平、焦书兰:《线索对基本特征刺激加工作用机制研究》,《心理科学》,2003, 26(5):814—817。
黄琨琨:《现代西方心理学十大学派》,安徽人民出版社 1990 年版。
黄希庭:《心理学实验指导》,人民教育出版社 1988 年版,第 1—39 页。
黄希庭:《心理学导论》,人民教育出版社 1991 年版,第 255—257 页。
金志成、何艳茹编著:《心理实验设计及其数据处理》,广东高等教育出版社 2002 年版,第 115—159 页。
李其维:《破解"智慧胚胎学"之谜——皮亚杰的发生认识论》,湖北教育出版社 1999 年版,第 122 页。
李新旺:《生理心理学导论》,河南大学出版社 1992 年版,第 108—163 页。
李亦菲、朱新民:《一种通用的口语报告编码方案》,《心理学动态》,1998(4)。
梁宁建:《当代认知心理学》,上海教育出版社 2003 年版,第 276—278 页。
廖伯琴:《中学生物理问题解决的表征差异及其成因探析》,西南师范大学博士学位论文,1999 年。
列宁:《列宁全集(第 14 卷)》,人民出版社 1985 年版。
林崇德、杨治良、黄希庭:《心理学大辞典》,上海教育出版社 2003 年版。
林崇德:《发展心理学》,人民教育出版社 1995 年版,第 77—88 页。
林崇德:《论儿童心理学与教育心理学研究的新趋势:生态学运动》,《心理发展与教育》,1990, 3:144—149。
刘婷:《生态心理学研究述评》,《东北大学学报(社会科学版)》,2002(2):83—85。

刘伟、袁修干:《人的视觉—眼动系统的研究》,《人类工效学》, 2000, (4):41—44。

刘艳芳:《S-R 相容性:概念、分类、理论假设及应用》,《心理科学》,1996, 19(2):105—109。

鲁忠义、杜建政:《记忆心理学》,人民教育出版社 2005 年版,第 25—27、38、47—48、231—233、252—265 页。

孟庆茂、常建华:《实验心理学》,北京师范大学出版社 2001 年版,第 33—35、47—49、107、140 页。

莫雷、李利、王瑞明:《熟练中英双语者跨语言长时重复启动效应》,《心理科学》,2005, 28(6):1288—1293。

欧阳虹:《面向 21 世纪的心理实验教学改革探索与实践》,《洛阳大学学报》,2001, 16(2):75—78。

彭聃龄:《汉语认知研究——15 年工作的回顾》,见中国心理学会:《当代中国心理学》,人民教育出版社 2001 年版,第 68—69 页。

彭聃龄:《普通心理学》,北京师范大学出版社 2001 年(第二版),第 125、143—144、152—153 页。

秦晓利:《面向生活世界的心理学探索——生态心理学的理论与实践》,吉林大学博士学位论文,2003 年。

阮宝湘、邵祥华:《工业设计人机工程》,机械工业出版社 2005 年版,第 32 页。

邵志芳:《思维心理学》,华东师范大学出版社 2001 年版,第 15、45—57 页。

沈德立、白学军:《实验儿童心理学》,安徽教育出版社 2004 年版,第 31—33 页。

沈德立:《关于儿童心理实验中控制因素的问题》,见全国高校儿童心理学研究汇编:《当前儿童心理学的进展》,北京师范大学出版社 1984 年版,第 161—173 页。

沈德立:《学生汉语阅读过程的眼动研究》,教育科学出版社 2001 年版,第 171、321—333 页。

舒华:《心理与教育研究中的多因素实验设计》,北京师范大学出版社 1994 年版,第 58 页。

宋广文、李寿欣:《场依存—独立性认知方式:理论演进及其应用研究》,《内蒙古师范大学学报》,1999(2)。

隋南:《脑、激素与记忆》,见中国心理学会:《当代中国心理学》,人民教育出版社

2001年版,第124—129页。

孙晔等:《出生后个体发展的心理生理学问题》,《心理科学通讯》,1982, 5。

汪安圣:《思维心理学》,华东师范大学出版社1992年版,第20、89—99页。

王美芳、张文新、林崇德:《6—12岁儿童嵌套思维的发展研究》,《心理科学》,2001, 24(4):492—493。

王甦等:《中国心理科学》,吉林教育出版社1997年版,第1—18页。

王重鸣:《心理学研究方法》,人民教育出版社1990年版,第55—56、104—105页。

吴凤岗:《青少年心理学》,北京师范大学出版社1991年版,第17页。

谢庆森、牛占文:《人机心理学》,中国建筑工业出版社2005年版。

徐嘉宏、吴英璋、余德慧:《台湾实验心理学的奠基者:郑发育教授》,《中华心理学刊》,41(2)。

许百华、傅小贞:《液晶显示的红、绿色光对视觉暗适应影响的比较研究》,《人类工效学》,2002, 8(4):17—19。

阎国利:《不同年级学生阅读科技文章的眼动过程研究》,《心理科学》,1999, 3:225—228。

阎国利:《眼动分析法在心理学研究中的应用》,天津教育出版社2004年版。

杨博民:《心理实验纲要》,北京大学出版社1989年版,第320、384—388、401页。

杨治良编著:《记忆心理学》,华东师范大学出版社1999年版,第9—10页。

杨治良:《基础实验心理学》,甘肃人民出版社1988年版,第2—7、90—148、186—187、218—223、256、288—293页。

杨治良:《记忆心理学》,华东师范大学出版社1994年版,第2页。

杨治良:《实验心理学》,浙江教育出版社1998年版,第4—32、140、186—187、235、468、598—600页。

叶浩生主编:《西方心理学的历史与体系》,人民教育出版社1998年版,第54—91页。

叶浩生:《心理学史》,高等教育出版社2005年版,第251—278页。

易芳:《生态心理学的理论审视》,南京师范大学博士学位论文,2004年。

余建英、何旭宏:《数据统计分析与SPSS应用》,人民邮电出版社2003年版,第141—159页。

俞文钊:《实验心理学》,浙江教育出版社 1989 年版,第 232、253、272—273 页。
虞积生、方俐洛、张嘉棠:《中国正常男青年的深度视觉阈限值的测定(Ⅰ)》,《心理学报》,1980, 12(3):298—302。
詹美莎:《北京市道路交通标志的心理学评价》,《心理学报》,1987, 19(2):167—174。
张厚粲主译:《国际心理学手册》,华东师范大学出版社 2002 年版,第 180—182 页。
张厚粲、吴正、宋华:《LOGO 程序设计教学对儿童元认知能力发展影响的研究》,《心理发展与教育》,1994, 1:1—5。
张厚粲等:《关于认知方式的测验研究》,《心理科学通讯》,1982(2)。
张侃:《刺激-反应相容性原理与工程心理学研究》,见中国心理学会:《当代中国心理学》,人民教育出版社 2001 年版,第 513—519 页。
张利燕:《心理学研究的生态学倾向》,《心理学动态》,1990(1):16—19。
张庆云:《普通实验心理学》,河南大学出版社,1993 年版,第 130、152—154、163—165 页。
张文新、纪林芹:《3—4 岁儿童攻击行为发展的追踪研究》,《心理科学》,2003, 26(1):49—52。
张学民、舒华编著:《实验心理学纲要》,北京师范大学出版社 2004 年版,第 1—23、189—190、201—202 页。

张一中:《心理学实验教学及其发展趋势》,《实验室研究与探索》,2002, 21(2):17—18。
张智勇、田立:《心理实验室建设与心理实验软件》,《心理学动态》,1996, 4(2):60—63。
中国心理学会:《心理学论文写作规范》,科学出版社,2001:88—91。
周海谦:《空间视敏度与时间视敏度的相关关系研究》,《许昌学院学报》,2004, 23(2):57—59。
周治金、陈永明、杨丽霞等:《汉语同音歧义词歧义消解的过程及其抑制机制》,《心理学报》,2003, 35(1):1—8。
朱新民、李亦菲:《架设人与计算机的桥梁》,湖北教育出版社 2000 年版,第 92—94 页。
朱滢:《启动效应——无意识的记忆》,见《当代心理学研究》,北京大学出版社

1993 年版。

朱滢:《实验心理学》,北京大学出版社 2000 年版,第 36—39、94—97、100—101、148—149、191、206—207、238—242 页。

朱祖祥、葛列众、张智君:《工程心理学》,人民教育出版社 2000 年版,第 151—155 页。

朱祖祥:《关于照明与视觉显示工效的若干研究》,见中国心理学会:《当代中国心理学》,人民教育出版社 2001 年版,第 482—484 页。

左斌:《西方社区心理学的发展及述评》,《心理学动态》,2001(1):71—76。

外文部分

Alluisi, E. A. & Warm, J. S. (1999). Things that go together, stimulus-response compatibility. In: R. W. Proctor (Ed.). *Stimulus-response compatibility: an integrated perspective.* North-Holland, Elsevier Science Publisher, BV, 33—86.

American Psychological Association(1995). *Publication manual of the American Psychological Association* (4th ed). Washington, D. C.

Anderson, J. R. (1980). *Cognitive psychology and its implication.* San Francisco: Freeman.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: a proposed system and its comtrol processes. In K. W. Spence(Ed.). *The psychology of learning and motivation: Advances in research and theory.* Vol 2, New York: Academic Press.

Atkinson, R. C., et al. (1988)(Eds.). *Stevens' handbook of experimental psychology.* New York: John Wiley & Sons, Inc, 205、1512.

Boneau, H. (1990). Psychological literacy: a first approximation. *American Psychologist*, 45:891—900.

Brown, A. L. & Barclay, C. R. (1976). The effects of training specific mnemonics on metamnemonic efficiency of retarded children. *Child Development*, 47: 71—80.

Brown, A. L. (1978). Knowing when, where, how to remember: A problem of metacognition. In: R. Glaster(Ed.). *Advances in instructional psychology*,

Hillsdale, HJ: Erlbaum. 367—460.

Brown, A. L., Campione, J. C. & Murphy, M. D. (1977). Maintenance and generalization of trained metamnemonic awareness in educable retarded children. *Journal of Experimental Child Psychology*, 24:191—211.

Bruner, J. S., Goodnow, J. J. & Austin, G. A. (1956). *A study of thinking*. New York, Wiley.

Cofer, C. N. (1967). Conditions for the use of verbal associations. *Psychological Bulletin*, 82:1—2.

Cooper, L. A. & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase(Ed.), *Visual information processing*. New York: Academic Press, 75—176.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11 (6):671—684.

Craik, F. I. M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104:268—294.

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 9:143—148.

Darwin, C. J., Turvey, M. T. & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: evidence for brief auditory storage. *Cognitive Psychology*, 3:255—267.

Ebbinghaus, H. (1885). *Memory: a contribution to experimental psychology*. New York: Dover.

Ericson, K. A. & Simon, H. A. (1984). *Protocol analysis: verbal reports as data*. Cambridge, MA: MTP Press.

Fitts, P. M. & Seeger, C. M. (1953). S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46: 193—210.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new of cognitive-development inquiry. *American Psychologist*, 34:906—911.

Fletcher, H. (1953). *Speech and hearing in communication*. New York: Van Nostrand.

Galotti, K. M. (1998). *Cognitive psychology in and out of the laboratory*. Washington, Brooks/Cole Publishing Company, p. 27.

Gavin, H. (1998). *The essence of cognitive psychology*. London: Prentice Hall Europe, 1998, 26.

Gibson, J. J. (1963) . *The useful dimensions of senditivity*.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Glanzer, M. & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 5:351—360.

Graf, P. & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11:501—518.

Graf, P. , Squire, L. R. & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10:164—178.

Greeno, J. G. (1978). The nature of problem solving abilities. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes (5): Human information processing*. Hillsdale: Lawrence Erlbaum Association.

Greenwald, A. G. , Pratkanis, A. R. , Leippe, M. R. , et al. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, 93:216—229.

Gregory, R. L. (1972). Seeing as thinking. *Times Literary Supplement*, 23 June.

Hockey, G. R. , Maclean, A. & Hamilton, P. (1981). *Attention and performance*. Vol, 9.

Holway, A. H. & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology*, 54:21—37.

Huey, E. B. (1980). The psychology and pedagory of reading. *The Macmilian Company*, 9:572.

Jacoby, L. L. (1983). Perceptual enhancement: persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9:21—38.

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30: 513—541.

James, W. (1890). *The principles of psychology*. New York: Holt.

Johnson, D. M. & Erneling, C. (1997). *The Future of the Cognitive Revolution*. Oxford University Press, Inc.

Just, M. A. & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indicates of sentence processing. *Canadian Journal of Experimental Psychology*, 47(2):310—339.

Kanizsa, Q. (1976). Subjective contoure. *Scientific American*, 234:48—52.

Kimble, G. A. (1967). Psychology's two culture. *American Psychologist*, 39(8).

Kinchia, R. A. & Wolfe, J. M. (1979). The order of visual processing: "top-down", "bottom-up", or "middle-out". *Perception & Psychophysics*, 25: 225—231.

Lange, G. & Pierce, S. (1992). Memory-strategy learning and maintenance in preschool children. *Developmental Psychology*, 15:543—551.

Lehrer R, Schauble L. & Petrosino A. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn & Tokada (Eds.), *Designing for science: implications from everyday, classroom, and professional settings*. Mahwah N J: Lawrence Erlbaum Associates: 251—278.

Levine, G. & S. Parkinson (1994). *Experimental methods in psychology*. New Jersey: Lawrence Erlbaum Associates, 1—3.

Lindsay, P. H. & Norman, D. A. (1977). *Human information processing* (2nd ed.). New York: Academic Press.

Lodico, M. C., Ghatala, E. S., et al. (1983). The effects of strategy monitoring training on children's selection of effective memory strategies. *Journal of Experimental Child Psychology*, 35, 263—277.

MacWhinney, B. , James, J. , Schunn, C. , Li, P. , et al. (2001). STEP-A system for teaching experimental psychology using E-Prime. *Behavior Research Methods, Instruments & Computers*, 33(2):287—296.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Mehler, J. , Bever, T. G. & Carey, P. (1967). What we look at when we read. *Perception & Psychophysics*, 2:213—218.

Michael, S. G. , Richard B. I. & George, R. M. (2002). *Cognitive neuroscience: the biology of the mind*. *Norton*, 184.

Milner, B. , Corkin, S. & Teuber, H. L. (1968). Further analysis of hippocampal amnesic syndrome: 14-year follow-up study of HM. *Neuropsychologia*, 6: 215—234.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64:482—488.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9:353—383.

Neisser, U. (1997). The future of cognitive science: an ecological analysis. In D. M. Johnson & C. Erneling (Eds.), *The future of the cognitive revolution*. Oxford University Press, Inc. , p. 247—260.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-century-Croft.

Newell, A. (1955). The chess machine: an example of dealing with a complex task by adaptation. *Proceedings of the Western Joint Computer Conference*.

Newell, A. & H. A. Simon (1972). *Human Problem Solving*. Englewood Cliffs: Prentice-Hall, 19—20、165—166、230—244、787—868.

Posner, M. I. , Boies, S. J. , Eichelman, W. H. & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 79(1, Pt. 2), 1—16.

Pressley, M. (1991). Can learning disabled children become good information processors? How can we find out? In L. Feagans, E. J. Short & L. Meltzer (Eds.), *Subtypes of learning disabilities: theoretical perspectives and research*. Hillsdale, NJ: Erlbaum, 137—162.

Reed, E. (1997). The cognitive revolution from an ecological point of view. In D. M. Johnson & C. Erneling (Eds.), *The future of the cognitive revolution*. Oxford University Press, Inc., p. 261—273.

Reicher, G. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81:275—280.

Reitman, W. (1965). *Cognition and thought*. New York: Wiley.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., et al(1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society American*, 37, 415—420.

Roediger, H. L., Weldon, M. S. & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: a processing account. In H. L. Roediger and F. I. M. Craik (Eds.), *Varieties of memory and consciousness: essays in honour of Endel Tulving*. Hillsdale, NJ: Erlbaum.

Scarborough, D. L., Gerard, L. & Corstese, C. (1979). Accessing lexical memory: the transfer of word repetition effects across task and modality. *Memory & Cognition*, 7:3—12.

Schiffman, H. R. (1996). *Sensation and perception*. John Wiley & Sons, Inc.

Shrauger, J. S. (1972). Self-esteem and reactions to being observed by others. *Journal of Personality & Social Psychology*, 23: 191—200.

Sidowski, J. (1988). The society: some history. *SCiP Newsletter*, 5:1—3.

Sperati, C. (2003). The inner working of dynamic visuospatial imagery as revealed by spontaneous eye movements. In J. Hyona, R. Radach & H. Deubel (Eds.), *The mind's eye: cognitive and applied of eye movement research*. Elsevier Science BV., 119—142.

Sperling, G. A. (1960). The information available in brief visual presentations. *Psychology Monographs*, 74:1—29.

Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta Psychologica*, 31:276—315.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64: 153—181.

Stevens, S. S. (1975). *Psychophysics: introduction to its perceptual, neural and*

social prospects. New York: Wiley.

Stevens, S. S., Morgan, C. T. & Volkman, J. (1941). Theory of neural quantum in the discrimination of loudness and pitch. *American Journal of Psychology*, 54:315—335.

Teasdale, J. D. & Fogarty, S. J. (1979). Defferential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, 88:248—257.

Tennant, M. (1988). *Psychology and adult learning*. London: Rontledge, 100.

Tulving, E. (1985). How many memory system are there? *American Psychologist*, 40:385—398.

Tulving, E. & Donaldson, W. (1972). *Organization of memory*. New York: Columbia University Press, 381—403.

Tulving, E. & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247:301—306.

Warren, R. M. & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American*, 223:30—36.

Warrington, E. K. & Weiskrantz, L. (1968). A new method of testing long-term retention with special reference to amnesic patients. *Nature*, 217:972—974.

Warrington, E. K. & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia*, 419—428.

Watson, J. B. (1914). *Behavior: an introduction to comparative psychology*. New York: Holt.

Waugh, N. C. & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72:89—104.

Wegel, R. L. & Lane, C. E. (1924). The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physical Review*, 23:266—285.

Weiskrantz, L. & Warrington, E. K. (1970). Verbal learning and retention by amnesic patients using partial information. *Psychonomic Science*, 210.

Weisstein, N. & Harris, C. S. (1974). Visual detection of line segments: an object-superiority effect. *Science*, 186:752—755.

Witherspoon, D. & Moscovitch, M. (1989). Stochastic independence between two implicit memory tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15:22—30.

Witkin, A. A., Moore, C. A. & Goodenough, D. R. (1977). Field dependent and field independent cognitive style and their educational implications. *Review of Educational research*, 47:1—64.

Woodworth, R. S. & Schlosberg, H. (1954). *Experimental psychology* (revised). New York: Henry Holt and Company.